GSA Office of Governmentwide Policy

U.S. General Services Administration • Office of Governmentwide Policy

NEWSLETTER

XML Applications in Government • Issue 11 • February 2002

This newsletter is available online at the Office of Intergovernmental Solutions Homepage at http://www.gsa.gov/intergov

Issues for Managers

By Frank McDonough
Deputy Associate Administrator
Office of Intergovernmental Solutions
U.S. General Services Administration

Introduction

The government of the future is being built today. Intergovernmental management, also called cross government management, is the foundation of the new form of government. Technology, especially Internet and Web enabled systems, encourages us to propose integrated solutions that cross jurisdictions and lead us into the future.

The critical question is how to achieve the integration. In the past, we thought the solution was to develop central databases formed out of back-end legacy systems. Little resulted from this approach because of the complexity of the task and resistance by the owners of the individual legacy systems. More recently, distributed databases and various middleware packages deal nicely with distributed heterogeneous data. However, the time to develop the solution often seems endless and the cost is quite high for the owners of the data. XML offers a better, lower-cost alternative.

A Brief Review of XML

XML is a family of technologies and standards. The basic specification of 1998 came from working group efforts of 1996. The second edition of the specification was published in 2000. Basically, XML is a method of structuring data that forms a common link between the many disparate and stovepipe systems that are now deployed by all developed government organizations in the world. XML is a tool that increases interoperability in cases where information transfer between disparate systems is important and where interoperability requires an improved process.

XML has become a universal format. It is a set of rules, guidelines and conventions for designing text formats in a way that produces Web-enabled files that are easy to produce and read.

Rapid Expansion of XML Use

XML is being used in widely in government and in at least 64 sectors in industry to integrate applications, connect government organizations and governments, manage content and knowledge, and accelerate cross organization collaboration.

The articles in this report from government and industry provide insights that enable us to confirm the rapidly emerging role of XML in connecting governments.

The articles in this report are nicely balanced, representing the U.S. government, state and local governments, international governments , and industry, generally reporting on activities performed under contract to the federal government.

All of the articles in this report discuss the consolidation of a government in some respect. The range of activities is impressive.

- Five of the initiatives document use of XML to consolidate government as a whole. Some have developed frameworks to achieve the consolidation. Governments active at this level include the national governments of Australia, France and the United Kingdom. In the United States, the states of Georgia and New York are taking a whole government approach as well.
- Six initiatives document the use of XML to selectively consolidate across government(s). These include:
  - The international coordination program for electronic filing of patents
  - The National Business Registry of five million firms in France
  - Implementation of the Geography Markup Language, a derivative of XML, in Australia, Germany, Japan and the United Kingdom
  - U.S. EPA's goal of a National Environment Information Exchange to allow sharing of data between states and the federal government
  - The National Institute of Justice's XML solution for cross-jurisdiction sharing of data
  - Wisconsin's Consolidated Court Automation to allow cross-jurisdiction sharing of data
- Eight governments are using XML to consolidate within functions such as law enforcement, finance, recreation, electronic bid solicitation,

In This Edition

U.S. General Services Administration
Office of Governmentwide Policy
Office of Intergovernmental Solutions

Co-editors
John Clark
James Mackison
Sally Matthews
Lisa Nelson

Comments and Suggestions
Your comments and suggestions are always welcome. Let us know if you find the information useful and what articles you like.

Send mail to:lisa.nelson@gsa.gov

Issues for Managers

national economic census, building sciences, national laboratory scientific proposals and translations of 14 foreign newspapers for intelligence purposes.

- Two initiatives are futuristic, including Voice XML, which can address the digital divide, and the development of a Human XML to include a person's motivation, intentions, attitude, psychological state, personal perspectives and beliefs.

Seven Special XML Issues for Senior Managers

Is XML the long awaited Holy Grail? Are there pitfalls as there were with so many silver bullets in the past history of information technology? Here are seven areas special to XML that senior managers should consider.

A clear picture is needed for when XML should be used.

- XML is not an across-the-board solution in IT. It is not a silver bullet. XML will not solve all of the problems in IT delivery. It is not an all-or-nothing solution.
- XML should be used in flagship programs, messaging and in databases and office products, when information transfer between disparate systems is important and where improvements are needed to allow interoperability. (See Molitoris for FAQ for managers.)
- The greatest benefit of XML will be realized in government organizations with multiple Web initiatives, multiple legacy repositories and a strict need for consistent and accurate content across these multiple sites. (See Hale.)

Are all of the needed standards available?

- XML's importance is based on its ability to describe products and standards that promote integration among diverse systems. However, it is the use of industry-accepted standards that allows XML to be successful as an integration and delivery technology. To date, a number of XML standards have been developed. More are needed. For example, an industry-wide standard is needed to describe processes or workflow so that governments can share those processes. Also, stronger security mechanisms are needed to ensure the confidentiality of constituent information. (see Miles.)

Will all XML investments pay off?

- Gartner Inc. conservatively estimates that, through 2004, organizations worldwide will waste at least $2 billion on XML specification activities with no return on their investment. Gartner believes that adopting *vocabulary-based standards* will provide the greatest benefits over the long run. (See Knox.)

In what ways should an XML investment pay off?

- XML should "future proof" information against periodic technology change, facilitate integration and promote collaboration.
- XML should reduce the cost of integrating data, replication of data and warehousing (where these are clearly needed).
- It should allow communication between applications running on different Web servers.

Where should XML data be stored?

- Managers need to consider whether they want to store XML data in a relational database designed for structured relational data. XML data is unstructured and hierarchical. Native XML databases should be considered, especially those with pattern recognition capabilities to address inter-governmental and intra-governmental requirements. (See Gurin.)

Managers need to create an environment for cross-sector solutions.

- The many efforts underway to develop XML data exchange specifications are often independent. These will lead to stovepipe solutions that may be inefficient and counterproductive. Senior managers need to take a broad view of XML activities and bring parties together to develop mutually agreed upon XML solutions that will further leverage the XML activities of separate but related groups. (See Linquist for an important reconciliation case study.)

What else should managers think about?

- Training is needed to increase awareness of the XML possibilities, spur collaboration and identify factors to be considered in making XML investment decisions.

Conclusions

XML is a family of standards and technologies. Since it formally arrived on the scene in 1998, a series of new capabilities have been added. Others, including security and query capabilities, are in the works.

As seen in this report, there are many and varied uses of XML in place or on the drawing boards in governments and other organizations around the world. The XML family of standards and technologies is helping to build the integrated government of the future.

Private Sector News

Digitizing the U.S. Census Bureau

By Steven A. Schafer, Ph.D.
CTO Fenestra Technologies Corporation

Modernizing the Economic Census

Our current work with the Census Bureau relates to a key initiative that takes place every five years, the Economic Census (see http://www.census.gov/epcd/www/econ2002.html). The Economic Census provides data on manufacturing, mining, retail and wholesale, trade, service industries, construction, and transportation. Millions of businesses are required by law to complete Economic Census survey forms.

Since 1993, Fenestra Technologies Corporation has helped transform Census business processes by leveraging numerous information technologies. The focus of our activities has been the redesign of paper forms and the creation of electronic versions of these forms. We have also developed systems to support:

- Paper and electronic data collection: Distribution of forms to respondents and collection of electronic response data
- Data capture: Recovery of response data from paper forms via optical character recognition (OCR) and key from image (KFI)
- Data processing: Analysis of response data and preparation of reports and other publications

To date, the bulk of this work has involved "unstructured" data data that are contained in various disparate and incompatible systems, without any sort of unified data model.

Shaping Dynamic Content to Increase Value

The infrastructure for the 2002 Economic Census consists of a central repository, containing both metadata, which define the data model and the characteristics of the forms distributed to respondents, and response data. Information is centrally located for ease of access and validation and consistency checking.

In addition to conventional paper forms, many respondents will be given the option to complete surveys electronically. Responses from electronic surveys will be automatically validated and entered directly into the repository.

Survey metadata, which define how a form is to appear to the user, are segmented into content, layout, and behavioral components. Content metadata control the actual wording of questions, instructions, etc. Layout metadata control the placement and typographical characteristics of text and images. Behavioral metadata (for electronic forms only) control the behavior of the form when buttons are clicked, inconsistent responses are entered, etc.

Fenestra has created four major software applications that work with the metadata in the central repository:

- Forms Designer: Allows non-specialists to design and lay out paper and electronic forms, using content metadata from the repository
- Autoformatter: Automatically assembles complete forms from stored layouts and content metadata
- Previewer: Allows designers to view the forms they have created
- Surveyor: Displays electronic forms to respondents and collects user responses

Index to Private Sector News

We integrated these applications with the repository using eXtensible Markup Language (XML; see http://www.w3.org/XML/) as the basis for information interchange.

XML Rationale and Considerations

A principal motivation behind the project is that Census wants to provide electronic forms based on exactly the same content as paper forms. This dictated the design of the central metadata repository. The choice of XML as the means of data interchange evolved out of our investigation of one of the XML family of standards, eXtensible Stylesheet Language (XSL), as a means of achieving the final form layouts (both electronic and paper). As it turned out, XSL was not up to the task at the time, so we developed our own variation of it. We did not use XFDL or XForms because neither supported the high-end typographical control required for the Census paper forms, qualities that are supported by XSL. Also, XFDL is proprietary.

We chose to use the same XML framework as XSL, rather than "reinvent the wheel." XML, an

accepted and stable standard, is well supported by various tools, so it was a choice of convenience as much as anything else.

XML is designed for interchange of data among disparate systems. Any large organization, such as Census, possesses a number of different computing systems that have grown and evolved over the years (and will continue to do so). Using XML as the basis for data interchange helps to insure that the information remains accessible and transportable as the systems change over time.

Key advantages of XML include:

- The XML vocabulary is sufficiently rich that it can be used to model essentially any kind of information that one might want.
- The simple, text-based syntax of XML eases data interchange.
- Tools for manipulating XML-formatted information are readily available, and XML interchange is supported by numerous applications.

There are also some disadvantages to XML:

- It is verbose, which can be a problem in some performance-sensitive applications.
- It contains a few arcane syntactical rules (primarily historical baggage) that can make it hard to dissect an XML-formatted document into its constituent parts.

For our purposes, we concluded that the pros clearly outweighed the cons.

Impact

To date, we have implemented paper versions of all of the 650+ forms (10–12 pages each) for the 2002 Economic Census, including those in Spanish. We are currently in the process of implementing electronic versions of these forms as well.

Census has integrated our software into its data capture process. Our system provides the answer-area coordinates of paper forms to the data capture software to assist with OCR and KFI subsystems. Also, behavioral metadata that our electronic forms use are used verbatim by the Census internal data capture system.

With these new systems in place, the Census Bureau will now be able to extract, analyze, and present data in a variety of forms across multiple publishing channels. We have enhanced their ability to provide quick access to accurate data that affect Congressional and Presidential policy decisions. Automation has also eased the burden for respondents.

Over time, the agency plans to build on this reservoir of knowledge, practical experience, and reusable software components to enable data collection and processing to proceed more accurately and less expensively in the future.

For more information, contact Steven A. Schafer on 301.916.8800, via e-mail at steve@fenstra.com.

What Will Make XML Usable?

By Rita E. Knox, Ph.D.
Vice President & Research Director
Gartner, Inc.

XML-based specifications are proliferating in every industry. While we understand the motivation for this, the proliferation is tantamount to proprietary. Even though the specifications are defined in XML, they use different element names, attributes, and data structures, and there is little hope that an independently developed specification in a similar or "adjacent" industry will be the same.

Foundations. The XML Family of Standards is more than just eXtensible Markup Language (XML) (see Figure 1). It consists of a host of standards, under development since 1997, that together create a foundation for Web-based computing.

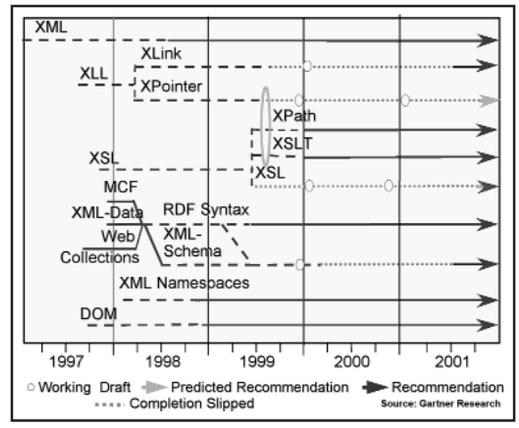


Figure 1. The XML Family of Standards (as of November 2001)

If the industry-specific standards created with XML are not interpreted in the same way by different processing applications, then the end results will not be the same. XML "Namespaces" is a specification for declaring which particular definition, for a component of a larger document type definition (DTD) or XML-schema, is being used. This is how XML parsers can determine if the XML model being used is actually the *same* model used by a particular enterprise. "Company" in one schema may represent the entity *sending* a purchase order; in another the same element name may represent a product manufacturer. "Namespaces" is the mechanism for the message originator to identify which definition is being used.

Framework. Our framework for classifying different XML constructs (see Figure 2) has five levels.

Private Sector News

- **Level 1:** The XML family of standards is like the materials used in manufacturing or the linguistic building blocks (characters and words) for language.
- **Level 2:** MathML, Scalable Vector Graphics (SVG) and so forth are specialized vocabularies for specific domains.
- **Level 3:** Universal Description, Discovery and Integration (UDDI), Simple Open Access Protocol (SOAP), and Web Services Description Language (WSDL) are XML-defined resource descriptions to provide information for enabling Web-based processes and are used by resource discovery processes. They correspond to linguistic grammars—the rules for stringing words together to create meaningful sentences—and manufacturing production processes.
- **Level 4:** Electronic Business XML (ebXML) and other standards are like parts of speech. They define a sequence of data structures required for a complete process-focused message, but focus on specific computing functions, such as process integration (e.g., "middleware") and business transactions.
- **Level 5:** Steel Markup Language (SML) and other XML-defined messages are complete data descriptions in a particular domain. They may be composed of discrete XML-defined terms (e.g., components written in SVG), but they are aggregates of those terms to support a complete process (e.g., a purchase order with buyer name, item, and delivery date required). The explosion of domain-, industry- and application-specific proposals illustrates that this is, by far, the most-prolific area of XML work.

Figure 2. XML-Based Building Blocks

Level	Metaphor 1-Language	Metaphor 2-Manufacturing	XML-based Construct
1	Alphabet, Word Formation Rules	**Materials Examples:** metal, plastic, lubricant, cable, rope, paint, leather, vinyl, cotton	XML Family of Standards
2	Dictionaries, Specialized Vocabularies	**Production Tools Examples:** wire harness, painting machine, upholstering, metal benders, drill presses, molds	**Examples:** CML. MathML, REDX,RNTD, SMIL, SVG, WebCGM, DII COE XS
3	Grammar	**Production Process Examples:** hire workers, assign jobs, define teams, define work stations, define sequence, define duration	**Examples:** SOAP, WSDL, UDDI, WSFL
4	Parts of Speech	**Production Steps Examples:** create automobile frame...add skin...install engine...install seating....upholster seating...paint...ship	**Examples:** ebXML, RosettaNet, Speech Synthesis, Speech Grammar
5	Sentences	**Finished Products Examples:** - purple car/leather interior - speed boat - twin-engine delta-wing jet - game console	**Examples:** SpacecraftML, SteelML, SPEC2000 (XML), MTML, PDM DTD, IXRetail

Source: Gartner Research

Activities. The Organization for the Advancement of Structured Information Standards (OASIS; host of http://www.xml.org/) has launched the development of Universal Business Language (UBL). UBL will provide a set of XML building blocks and a framework that will enable trading partners to exchange business documents in specific contexts.

Enterprises eager for standard, reusable vocabularies to facilitate e-commerce no longer fret over who will win the markup language skirmish. Rather, they want to know how to:

- Build transactions in XML
- Exchange data with business partners, customers, and vendors

Private Sector News

- Assure longevity in the XML data messages that they do adopt

Instead of a "one true way" consensus XML language, Gartner believes that the answer lies in building e-commerce transactions from *reusable* XML vocabularies that adopt the same terms when the same meanings and interpretations (by applications) are intended. The components of business exchanges (e.g., "company," "address," "part number," and "price") recur in different messages and among different industries. However, custom XML-defined transactions in every line of business would perpetuate proprietary solutions. If each exchange is unique (i.e., it uses new terms even if the same meaning is intended), computers will confront a tangle of incompatible languages; therefore, they will be unable to share data or reuse software to process the same data stream. OASIS knows about the proliferation of markup languages it maintains a database of proposed specifications (see www.xml.org). By taking a mediating role in resolving the markup language fracas, OASIS can help achieve cost and time savings in many processes in many industries.

Initially, UBL aims to prevent chaos by defining a specific vocabulary for e-commerce. If successful, UBL may then prove useful as a framework for creating "words" in other fields such as science, healthcare, politics, or art. Despite uncertainty over how this initiative will be executed, OASIS's attempt to create vocabularies rather than complete transactions takes a step that enterprises understand and seem ready to support.

Major initiatives for addressing these naming issues have been initiated by the U.S. Department of Defense:

- The Defense Information Systems Agency (DISA) has worked on the Common Operating Environment (COE) for at least four years. They maintain an XML registry in the COE Data Emporium that maintains data elements for multi-service reuse (see http://diides.ncr.disa.mil/xmlreg/user/index.cfm). It is designed to be a namespace for registration and reuse of common elements to support interoperability across the services.
- The Navy has launched Task Force Web (see http://ucso2.hq.navy.mil/n09w) to implement a "Webified" Navy within the next five years. XML is at the heart of the initiative; a Department of Navy (DON) XML Work Group is defining the framework for creating a robust XML infrastructure.

Conclusion. Through 2004, enterprises worldwide will waste at least $2 billion (estimating conservatively) on XML specification activities with no return on their investment. Industries and enterprises that adopt vocabulary-based standards development, however, will reap the greatest benefits over the long term--robustness, reusability, cross-industry compatibility, and lower development costs.

For more information, contact Rita Knox via e-mail at rita.knox@gartner.com.

XML's Role In The Geospatial Information Revolution

By Mark E. Reichardt
Director, Marketing and
Public Sector Programs
OGC, Inc

Virtually all sectors of government at all levels depend on geographic information. Everything and everyone is somewhere, and everything happens somewhere, so geospatial information ("geodata") is, or should be, a critical part of government information systems.

Unfortunately, easy sharing of geodata has been prevented by:

- Lack of open interfaces and protocols that enable interoperability between geoprocessing systems from different vendors and also between different kinds of systems—geoprocessing systems such as GIS, earth imaging, and navigation, and also otherwise non-spatial systems such as database software and CAD software.
- Lack of common naming schemes for geographic features and lack of common schemas for spatial metadata—descriptive data about geospatial data sets.

But those obstacles are finally being overcome. Global geoprocessing industry vendors and major government and industry users are creating open interfaces and protocols through their participation in the Open GIS Consortium, Inc. (OGC) (http://www.opengis.org). And data coordination groups around the country and the world, notably the Federal Geographic Data Commit-

tee (http://www.fgdc.gov), have made good progress on geodata semantic standards.

XML provides part of the solution to both obstacles. XML encoding of geodata using OGC's GML standard, as described below, makes it possible to display, overlay, and analyze geodata on any Web browser, even if the browser is getting views of different map layers from different remote map servers. The same data can be presented in countless different ways to suit different applications and devices. And XML is helping in the area of naming schemes and metadata schemas. XML and GML are easy grist for software that searches and parses and tries to make sense of different text-based metadata files that were created by different groups at different times for different purposes.

This progress is timely. Cell phones and other digital devices are rapidly gaining location awareness and Web connection, promising new spatial technology applications that will yield vast amounts of spatial information. The "spatial Web" is advancing rapidly, but the degree to which this progress will support e-government depends on governments' commitment to the standards initiatives described above.

Other countries, including Australia, Germany, Japan, and the United Kingdom, are moving quickly to implement the standards and knowledge they have helped to develop in OGC. For example, Ordnance Survey, the United Kingdom's national mapping agency, has standardized on GML as their geodata distribution format. They call their program "OS MasterMap". An article in the August 2001, issue of *GeoEurope* describes the program under an earlier name, Digital National Framework (DNF): http://www.geoplace.com/ge/2001/0801/0801gml.asp. A direct link to the Survey's OS MasterMap site is http://www.ordnancesurvey.co.uk/os_mastermap/home/home.htm. The MasterMap technical document is available for downloading as http://www.ordnancesurvey.co.uk/downloads/mm/osmm_in_gml_format.PDF.

In the United States, the Census Bureau is moving rapidly to take advantage of GML, as described below in a column which appeared in the May 2001, issue of GeoInformatics, a European geoprocessing industry magazine.

OGC invites agencies at all levels of government to participate in its active Technical Committee process and in its dynamic series of Interoperability Initiatives—testbeds, pilot projects, technology insertion projects, planning studies, and feasibility studies.

For more information, contact Mark Reichardt by telephone at (301) 840-1361 or via e-mail at mreichardt@opengis.org

Government's Case for OpenGIS - *A U.S. Census GML Example*

Louis Hecht,
Open GIS Consortium, Inc.

Why do Government agencies participate in OGC? Consider the U.S. Census Bureau. "Census" has been an OGC (Open GIS Consortium) member for over three years. Recently, Census's Technical Committee (TC) representative, Paul Daisey, has been most involved with the TC's effort to define Geography Markup Language (GML), an eXtensible Markup Language (XML) encoding of OGC's OpenGIS Simple Features.

The TC's GML Special Interest Group (SIG) published a recommendation paper for GML Version 1 early in 2000. Web Feature Servers (WFS) that serve GML to Internet clients and a draft WFS specification were developed in the latest OGC Interoperability Program Testbed.

Since the 1980s, Census has used its own data format called TIGER/LINE. Recently, Census hired a contractor to build an open source Java TIGER/LINE-to-GML translator, and Census intends to use a WFS later this year to serve TigerGML as a proof of concept prototype.

The Geography Division of Census maintains the TIGER database of Census geography to geo-locate all residential addresses in the U.S. and to support automated generation of 20 million map sheets for conducting and publicizing results of the Decennial census and the ongoing American Community Survey. In addition to using housing unit information collected by field staff and extracted from imagery to keep TIGER up to date, Census needs to exchange data with State, County, Municipal and Local governments across the country. Some of this exchange is conducted using commercial GIS tools. Some of it is done by custom-coding interfaces between their home-grown systems and Census's. Much of it is still done by digitizing paper maps.

TIGER is based on a Census-developed network database updated via Census-developed software for updating and digitizing. TIGER database software and applications were very successful innovations when they were developed in the late 1980's, and subsequent enhancements helped ensure that the U.S.

Private Sector News

2000 Decennial Census was the best ever. For that census, field staff went out with paper maps and pencils to identify housing units, and programmers enhanced the Master Address File (MAF) and TIGER databases and software. Census recently began a modernization effort to replace and enhance MAF/TIGER with commercial off-the-shelf (COTS) products, to extract features from imagery, and to provide mobile digital support devices for field staff.

OpenGIS services that can interoperate using GML over the Internet will allow Census to:

-- Distribute TIGER/GML using Web Feature Servers instead of shipping paper storage forms or CD/ROMs in proprietary data formats

-- Begin Web-based transactions between WFS clients and servers allowing the possibility of live updates from their partners

-- Buy Standards-Based COTS (SCOTS) products instead of building home-grown Census applications, so that Census will have "best of breed" components that "plug and play" over standard interfaces.

By the end of 2001, the TC hopes to have completed XML Version 3, which will include two components of special interest to Census:

1) GML extensions to the "OpenGIS Spatial Locator," a geocoding services standard developed in the Geospatial Fusion Services part of the IP2000 Web Mapping Testbed Phase 2.

2) A GML Point of Interest / Area of Interest standard: POI/AOI is a Spatial Locator, plus a direction, plus some digital content. Example: a point on a street corner in Boston, a compass heading, a digital photo of the department store on the opposite street corner, and a movie clip of crowds of people crossing the street. It could encode the output of a Cell Phone / GPS / Laser Range Finder / Digital Compass / Digital Camera device held by a Census worker identifying housing units.

OGC's OpenGIS Specifications help Census to plan their information infrastructure future with considerable certainty, and to plan and budget geospatial services that will be used across the range of information environments that support their workflow, from mobile wireless to wired web browser to geospatial database server. Widespread adoption and implementation of such standards in commercial products will allow large data set managers like Census and Location Based Service providers to buy more capable and less expensive tools, to share data, and to deliver data and services to much larger markets.

XML: Fast Tracking Content Solutions

How Content Infrastructure Harnesses XML to Increase Efficiency

By Mark A. Hale, Ph.D., P.E.

It is well understood that government agencies need the ability to produce accurate content for consumers, inter-agency interaction, and internal efficiencies on a routine basis.

XML has gained a tremendous amount of traction in recent years. It successes are due in part to its simple nature – it can be used shore up leaks in the back-end infrastructure. Once these holes are plugged, content flows richly from user to the destination site.

As a content infrastructure and a technology leader in the standards community, Interwoven is often looked to for a direction for a starting point for XML. There are a number of domains that are applicable. However, two stand out as promising for those still planning an XML rollout.

Templating on the Fast Track

Templating harnesses the power of XML to separate form from content. Solutions that minimize risk and have the largest return on investment are those that use XML to capture and manipulate patterned content – the kind of content that can be captured in forms. For instance, the form data that could be captured electronically by mandate of forms available under the Government Paperwork Elimination Act.

The National Library of Medicine is a classical example of maximizing process efficiencies. The Library collects materials in all areas of biomedicine and health care, as well as works on biomedical aspects of technology, the humanities, and the physical, life, and social sciences.

With a collections of more than 5.9 million items – books, journals, technical reports, manuscripts, microfilms, photographs and images, it was imperative for NLM to implement a content infrastructure solution that would enable effective search for their Web site and to enforce a standard look-and-feel all throughout the site.

NLM uses Interwoven®, TeamSite®, Templating to further leverage the power of XML. Lets examine what capabilities are exploited to permit NLM to maintain an XML solution that enables content re-use across multiple devices, brands and channels.

- User forms are coded in XML. This permits straightforward cre-

ation of new forms and customizations for specific initiatives. These forms are eventually rendered as Web pages for data capture.

- Data records are captured as XML-formatted files. This permits the content to easily be reused in multiple locations, for instance on several Web pages at once. In turn, this content can easily be indexed for search or even metadata can be captured at the same time as the newly entered content.
- A transformation engine is built into the solution. XML undergoes frequent changes and manipulation as it makes its way to Web pages and other output formats. The integration of stylesheets should not be an ad hoc activity.

In summary, Terry Luedtke, Computer Specialist at National Library of Medicine, says: "Content reuse enables more efficient use of resources such as people, time and existing content."

Content Solutions

Templating standardizes interfaces and that are routinely used throughout the enterprise. The resulting data can quickly be put unto the Web, sent in an e-mail, or archived in a library.

However, collecting content in XML is really the first step at a comprehensive solution. For instance, XML, stylesheets, formatting objects, and graphics can be combined to make material available for publication.

Interwoven TeamXML is a product offering that extends templating to bring full XML solutions to the mainstream. What does this mean to the bottom line?

- Parametric search is available on the objects. That means users can search against specific fields in their content, for instance "select all articles in which cardiology appears in the title."
- XML components can be assembled into catalogues and libraries.
- A full technical publications solution is available from partners including ArborText, SoftQuad, DeskNet, XyEnteprise, and SoftwareAG. This permits not only XML content entry, but layout and proofs.
- Legacy content can be converted to XML. This interface permits Word, PDF, PowerPoint, and more than a hundred other formats to be used in conjunction with patterned records from templating.
- XML allows content to be componentized. This means that components of an original document can be reused in multiple documents. For instance, a government disclaimer can be added to each XML document without having to re-write it every time.
- All content infrastructure functionality applies to each and every XML asset. This includes benefits like tracking of changes, approval via workflow and archive capabilities.

These are just a few ways in which to extend the base capability of XML content to extract even greater returns and efficiencies.

Summary

XML should not be adopted because it is the latest and greatest technology trend. Government Agencies should harness XML where and when they have the greatest need. The greatest benefit will be realized in organizations with multiple Web initiatives, multiple legacy repositories and a strict need for consistent and accurate content (in the form of XML components) across these multiple sites.

Beware approaches that demand all or nothing adoption of XML. This is not only overkill, it will also ultimately be cost prohibitive.

For more information, contact Dr. Hale at 408.220.7483, via e-mail at mark.hale@interwoven.com or visit the Website, www.interwoven.com

Achieving True Information Freedom: *Native XML Databases and Pattern Processing*

By Kathi Gurin
NeoCore Inc.

Significant heartache continues over achieving true value from the vast amounts of information held across government agencies. While we've all gone boldly into the "information age," the very technologies that enabled us are now constraining us from taking advantage of our most valuable asset information.

The advent of eXtensible Markup Language, or XML, brings new hope for information management. XML encompasses both data and information about the data (the metadata) and creates the opportunity to share rich, dynamic data more easily. However, with the benefits of XML also comes a new technical challenge—storing this new, unstructured data format.

This article is intended to introduce you to database technologies that address the efficient storage of and access to XML information. Storing XML natively is imperative. So is processing this information through pattern recognition, rather than with

traditional textual processes. This approach is necessary to overcome the otherwise significant performance and storage inefficiencies of working with XML.

An Information Wish List

I'm willing to bet that every government agency would desire the same high-level wish list for the use and management of information:

- Intra- and Inter-agency sharing of data
- Custom, secure views (often of the same data) for different users
- Aggregated or whole view of data residing on stovepipe systems today
- Information "agility"
- Ability to easily adapt a database for changed and new requirements
- Ability to easily support varied, heterogeneous data types
- Reduced costs related to information management
- Increased operational efficiencies

Certainly in today's economic and political climate, it is more critical than ever to be able to leverage all information, regardless of where it resides or in what format it exists. Data management has already evolved considerably in an effort to support business requirements—moving from application silos with no data integration to data being integrated across entire value chains. And, XML has greatly eased companies' data integration burdens and become the transport and integration platform of choice. But to truly address the "wish list" requires more than the ability to move dynamic data across system. What is needed is a robust, information-centric, XML storage and access mechanism without inherent (negative) performance and storage impacts.

Impracticalities of RDBMS with XML

A natural inclination of those embracing XML is to attempt to store XML data in a relational database. The problem with this is that RDBMSs were designed for structured, relational data. XML, being unstructured and hierarchical in nature, is anything but that. Forcing XML into a relational structure inhibits the use of its most powerful characteristics. XML self-contains the definitions of data and relationships between data and is inherently extensible. To store it in an RDBMS requires that you either store documents as CLOBs or BLOBs, where all functions are done at the document level, or you separate the data from the metadata, or "shred" it, index each component, and then map it into the relational structure. Access of the data in its original XML form requires re-assembly of the data with its metadata. The cost of shredding, indexing, mapping, and re-assembling the data using traditional methods is significant in both performance and storage. For example, storage of an XML document in an RDBMS can result in a footprint of up to 25 times the size of the source data.

Another compelling argument against storing XML in an RDBMS is the constraint that is put on the use and adaptability of the data. With an RDBMS, one must predetermine the use and users of the data and create a detailed design, or schema, for the data in order to store it. The likelihood of correctly forecasting the future requirements for the data is slim, of course, and subsequent changes to the schema are time consuming and very costly. Application lifecycle costs sky rocket, and projects are often over-budget, delayed, and often never even get off the ground because of the database implications. The ideal goal, of course, when using XML, is to use a storage method where extensibility is not lost. After all, why not take advantage of all that XML offers?

New Tools for Dynamic Information Management

To achieve a truly extensible database foundation that leverages the flexibility of XML, consider a native XML database. By definition, an XML database is an information store that manages XML natively. By supporting operations on both the data and metadata, an XML database handles information as a whole, understanding the relationships between content, context, and structure. XML databases are ideal solutions for applications that need to store and manage ever-changing data and metadata. Dirk Coburn, research director at IDC, says "This is the year when XML-based technologies and applications become mainstream... That transition is driving the growing need for robust solutions to manage XML. What IDC calls 'XML and virtual DBMS' software is central to this need and should be considered by enterprises anticipating a significant scope of XML-based e-business functionality when performance, ease of extensibility, and cost are of concern. It's one thing to be XML-enabled, but more than incidental use requires a higher level of functionality. Today, only native XML databases offer full XML information management."

Of course, even native XML databases come in different flavors. There are **Document-Centric** solutions where all database functions are performed at the document level; **Element-Centric** solutions, where individual elements within the

Private Sector News

document can also be searched and manipulated; and **Pattern-Centric** solutions, which again provide element level granularity, but also ensure significantly improved performance and a much smaller storage footprint. While **Document-Centric** methods are satisfactory for content management applications operating on small documents or those not requiring high performance, **Element-Centric** methods are best for applications with heavy transaction volumes and/or large volumes of structured XML. Recently available **Pattern-Centric** XML database solutions are unique in their ability to support element-level functions, while also satisfying the highest-level performance requirements. Pattern processing uses numeric processes against icon representations of the data (instead of indexing and textual-based searches) and achieves scalability of queries through near-flat performance, despite the size of the XML dataset. This is unheard of in the RDBMS world.

Summary

With a continuing growth in demand for collaboration, shared data, and flexible, agile systems comes new technical opportunities. A new data definition model, the eXtensible Markup Language, or XML, is rapidly being adopted due to its promise of extensibility and ease of data transport across systems. And now a new kind of information management and processing method has formed that is designed specifically around XML. Breakthrough, high-performance native XML databases that truly support the extensibility and self-describing nature of XML are soon going to give RDBMSs significant competition. Native XML databases, especially those with inherent pattern processing capabilities, are well positioned to address the inter- and intra-governmental information needs of today.

For more information, contact Kathi Gurin via e-mail at kgurin@neocore.com.

HumanMarkup Language As Data, and Within Data

By Ranjeeth Kumar Thunga
Chair, OASIS HumanMarkup Technical Committee

As a society and within our government, we have celebrated having immediate access to data. Recent technical advances have enabled us to transmit and exchange information instantly and reliably. XML standards are starting to become very promising as universal formats of data representation and exchange. We may conclude that these proposed solutions will finally eradicate our data communication problems. Yet, something essential still appears to be missing from our systems.

Systems based on current XML formats are able to deliver data in a manner all platforms can readily interpret, transfer, and store. Yet, when all is said and done, we must sit back and answer the following questions as well: Is the data clear, not only in its content, but also in its context? Does this data adequately represent the individuals it is supposed to? Is this data meaningful for you, for them, for us? And, does this data truly reflect a system of government "of the people, by the people, and for the people?"

Very often, the answer is "no.". Fundamental human characteristics within data are often ignored, abandoned, or even deliberately removed. Thus, Human Markup Language (HumanML) has been launched as an XML specification designed to represent and embed human information explicitly as data and metadata. What is this human information? This is any characteristic that uniquely describes aspects of us as human beings. This can include our basic physical characteristics, as well as our intentions, motivations, socio-cultural descriptors, beliefs, values, and behaviors. Further, constructs representing our mindsets or psychological schemas can be made explicit through a combination of several different HumanML modules.

Anecdotal summaries, which have so far been the only means of qualitatively representing human characteristics, cannot be interpreted or parsed by applications. Thus, it is now up to us to design systems that can make human information available as explicit, parsable data. Data representation of human characteristics would have to be fluid, flexible, and contextual, yet at the same time adhere to universal standards. All these needs can be met through XML. Thus, by designing an XML framework of human characteristics, we can greatly clarify the scope, purpose, and context of information by effectively representing human information as data.

There are numerous ways in which HumanML can improve current systems. An obvious example is in the enhancement of electronic records management. In current systems, a minimal set of human information is described. This may include the name of the person, appropriate contact information, and perhaps geographical address. However, these systems may be neglecting other human characteristics that may be absolutely essential for proper interpretation. These charac-

Private Sector News

teristics could include the individual's background, motivation, intention, attitude, psychological state, personal perspective, and beliefs. Although these human details may never have been assembled before as explicit data, they may very well be essential in interpreting records within their proper context.

Human information may be used not only as metadata, which provides context for pre-existing data, but also as data itself. We can adapt current systems, which already have the function of representing human data, towards XML-based representations of the same data. For example, public safety and social services databases, which currently may store human information, could soon adopt a universal format for exchanging data based on XML, through HumanML. This would allow far more precise descriptions of individual characteristics and far more efficient exchange of data between one another. Culturally and contextually, specific frameworks can be easily customized and elaborated on to the degree they are needed, in an explicit fashion, through HumanML.

When these systems are developed, they would have to inevitably address some important concerns—namely, the privacy, invasiveness, and security of such information. It would be up to each application and agency to determine how to best gather, represent, and utilize this data in a manner that is representative and helpful, without being obtrusive. We are continuing to explore a means by which we can provide individual access to a unified repository of human data, which can then be shared with appropriate agencies.

An example of using a HumanML-based data repository could be in allocating resources to individuals for various types of government assistance, including disaster relief. Unique human considerations need to be taken into account in all solutions provided. For example, during tragedies such as the September 11th terrorist attacks, custom-tailored action plans could be automatically generated, based on a HumanML representation of the cognitive framework of individuals. For refugee relief or foreign aid, cultural context and individual specifics could be embedded directly within data. This would allow us to accurately develop solutions that automatically take these factors into account. Ultimately, a system of describing human characteristics through XML could allow for far more precise, targeted appropriation of government efforts than either impersonal statistical information or anecdotal summaries could provide on their own.

How can such a system be technically developed? Since HumanML is in fact XML, we can use all of the related XML technologies to help build this infrastructure. Current XML-based technologies can be used to transform (XSLT), message (SOAP), present (XSLFO), query (XQuery), validate (XML Schema), and annotate (RDF) this information.

Our effort, which began in early 2001, is being sponsored by OASIS as an official HumanMarkup Technical Committee (http://www.oasis-open.org/committees/humanmarkup). OASIS is an international Internet consortium working with several hundred individual, corporate, government, and NGO members, as well as the United Nations, in developing universal non-proprietary open standards of information exchange through XML. Additionally, we have recently organized HumanMarkup.org, Inc. as a separate non-profit company designed to support this work, by developing solutions and applications based on these specifications.

We are currently in Phase 1 of the effort, which involves identifying the requirements and frameworks we will represent within HumanML. This is an open source effort, and everyone is invited to participate in the discussion. Additionally, we are hoping to develop formal relationships with government agencies interested in implementing these solutions, by gathering information about their current needs. HumanMarkup.org, Inc. has been designed exclusively for this function. Please contact us if you or your agency is interested in exploring these solutions. With human specifications in place, we can finally ensure that our government systems truly represents "We the People."

For more information, contact Ranjeeth Kumar Thunga, Chair, OASIS HumanMarkup Technical Committee and President, HumanMarkup.org, Inc. by telephone at (646) 456-9076 or via e-mail at rkthunga@humanmarkup.org.

State and Local News

XML in Government "Smart Portals"

By Moses E. Miles III
Chief Architect
State of Georgia

Faced with the challenge of shrinking budgets and a demand for higher levels of service from an increasing population, governments around the world are leveraging information technology to meet these challenges. Most strategies focus on providing alternative channels that allow governments to deliver information and services to their constituents. These channels include telephone, Internet, and wireless access. The platform that provides these alternative channels is now termed "smart portal."

Smart portals are customer-focused information delivery systems that provide users with consistent and reliable access to information services. The smart portal "virtually integrates" governments, by providing constituents with seamless access to information services irrespective of a government's organizational structure. Virtual integration provides opportunities for government entities to identify and eliminate duplicate services and to re-package complimentary service offerings from different government organizations into a single comprehensive offering tailored to the constituent.

The eXtensible Markup Language (XML) is the cornerstone enabling technology for smart portals. The importance of XML in smart portals is directly related to its flexibility and portability. The XML language has evolved into an industry standard for application and data integration. Standards such as Simple Object Access Protocol (SOAP) and Universal Description, Discovery Integration (UDDI) allow applications to share data and functionality independent of the programming languages used to build those applications. Equally important are XML schemas, which are data standards defined in XML that describe a common syntax for sharing data between information systems. Examples include LegalML, a standard that describes how to share electronic versions of legal documents.

The State of Georgia has adopted XML as the enabling technology for its smart portal. The state is in the process of constructing its Enterprise Portal and Interoperability Architecture (EPIA) that will serve as its sole electronic information and service delivery platform. The EPIA was developed under the following architectural requirements:

1. Existing information technology systems are valued resources that must be leveraged well into the future.
2. Any State-of-Georgia government entity that participates in the portal will require autonomy over its data.
3. Integration must be platform independent.
4. The system must support telephones, Web, cell phones, personal digital assistants (PDA), WebTV and any future device as information delivery platforms.
5. Users must be able to access services based on their intentions.

Given these broad requirements, the state developed a service-oriented architectural framework composed of services and customer-centric business processes. This framework consists of four major components, including service providers, connectors, middleware, and display devices.

Service providers are existing and future systems that present application functionality to the middleware in the form of a service. Connectors provide the translation mechanism between disparate technologies. For example, an application developed in the archaic Total programming language of the 1970's could present the middleware with an XML-based interface to its business logic through its Service Provider Connector. Likewise, the Interface Connector for an Interactive Voice Response (IVR) system could accept a XML document from the middleware and transform the data into speech for a telephone. The Middleware is composed of a variety of components designed to manage the user's interaction, services, and processes. Components of the middleware include customer relationship management (CRM) software, content management software, and service and process repositories. Lastly, display devices are the presentation technologies the architecture supports. (For more information about the State of Georgia's architecture see www.gagta.com.)

Within this framework, services are defined as distinct pieces of functionality or content delivered by an information system. Examples include credit card processing, change of address, and the driver's license manual. Processes are similarly defined as the aggregation of one or more services with execution

Index to State and Local News

logic. Processes are further required to be customer centric. Users are expected to interact with these processes rather than services because of the unmanageability of the estimated 500,000 services provided by state information systems today. Aggregating these services based on user intentions should fundamentally improve the way constituents interact with governments within the State of Georgia.

The two XML technologies that the State of Georgia EPIA will use to enact its architectural vision are Web services and XML schemas. A Web service is multi-layered protocol used to describe functionality presented by an application. The Web services protocols include HTTP for transport, XML for data, SOAP for routing, UDDI for Discovery, and WSDL (Web Service Description Language) for description. The back-end applications of the EPIA will provide services as Web services through connectors as necessary. These Web services will reside in a UDDI repository inside the middleware. Once registered inside the UDDI repository, these services will be aggregated to form processes within a tool the state refers to as a process manager. This process manager will build XML-based processes described by languages such as WSFL (IBM's Web Service Flow Language) or XLANG (Microsoft's equivalent). These languages describe workflow of service execution, including the order of execution and workflow logic. The data in these services or processes will be sent to the presentation layer of the middleware as an XML data structure that will be mated to an XML style sheet (XSL) or transformation (XLST) that will generate the presentation for the targeted device, such as VoiceML for telephones or wireless application protocol for cell phones.

In addition to Web services, the EPIA will make extensive use of XML schemas. The state will build a comprehensive data dictionary using XML schemas. The schemas will form a hierarchical data structure allowing data to be encapsulated and passed between systems of the same class. The hierarchical nature of the data dictionary will be extensible and adaptable enough to allow state entities to continue to build information systems in a silo manner, but still benefit from the integration provided by the EPIA. Lastly, by having a common data dictionary/structure, the government entities of the state can store data with autonomy and transmit that data as an XML structure that adheres to the state's data structure.

XML is a crucial technology for smart portals by providing universally accepted mechanisms for application and data integration. XML's importance is based on its ability to describe protocols and standards that promote integration among diverse systems. However, XML is only successful as an integration and delivery technology with accepted industry standards. Although the industry has developed a number of XML-based standards, today an industry-wide standard is needed to describe processes or workflow so that governments can share those processes. Finally, the Web services protocols will require stronger security mechanisms to ensure the confidentiality of constituent information in smart portals.

For more information, contact Moses E. Miles by telephone at (404) 463-2310 or via e-mail at mmiles@gta.ga.gov

Extending New York State's Digital "Government Without Walls" Through XML

By William F. Pelgrin,
Commissioner,
New York State Office for Technology

On March 1, 2001, Governor George E. Pataki announced the official release of New York State's new Portal Home Page (http://www.state.ny.us) and the state's Common Web Banner that is on all executive agency home pages. This banner (see below) provides easy and convenient access to government information, services, and transactions available on the Internet from a central location.

This release marks the first major milestone in Governor Pataki's e-commerce/e-government initiative, which he officially launched on June 12, 2000. The theme of this initiative is "Government Without Walls." It is intended to break down barriers for citizens doing business with individual state agencies, allowing citizens to conduct business and find information anytime, anywhere without having to know which agency handles each type of transaction—in essence, a one-stop shopping concept.

Figure 1: New York State's Web Banner shows the various landmarks of the State that includes a sailboat, the Statue of Liberty, New York City Skyline with beams of lights representing the World Trade Center building with an inscription "In Remembrance". The banner also has four navigational boxes to various government services.

By navigating through the portal, users can find information in a variety of ways. The user is able to browse categories of life events and find links to both appropriate state agency Web site pages and relevant Frequently Asked Questions (FAQs). A search feature is also available to find the same information based on user-supplied keywords. All links and FAQs are stored in a central Web-accessible database that more than 62 agencies use to keep information, services, transactions content of the portal, and Web banner buttons up-to-date. Additionally, the state is currently piloting the participation of local governments within the portal. New York City now enters links and FAQs into the central database so citizens can easily find information and services for both the city and the state.

As the New York State portal continues to evolve, there has been more effort devoted to integrating information from different state agencies, so that users will not have to go to another Web site to find the information they need. The Office for Technology, however, does not create or maintain most of the information available through the site. In order to make current, accurate information available, syndication is being employed to retrieve the requested data directly from the source agency and re-purpose the content so that it has the same look and feel as the rest of the portal. Extensible Markup Language (XML)—a universal format for exchanging structured documents and data on the Web—is being used to package the information as it is transferred between Web sites.

New York State is currently testing syndication for use within the state portal and anticipates being production ready by the first quarter of 2002. Syndication is being implemented for three different sets of data in the portal. One example is the use of a geographic (map) interface to navigate to a county of interest. On the county page, which can be accessed from the map-NY button on the common Web banner, there is a list of county officials with contact information. This information is retrieved from another state agency that maintains an accurate list of local officials to meet its own mission. Although data is cached on a daily basis for performance reasons, the information presented to the user is coming from the most reliable state source and is always current. In addition, a list of municipalities within each county is also available on the county page from the same source. A third example where syndication is being implemented is the search feature enabling citizens to look up their legislators or election districts. The information is retrieved using XML and presented to the user in the same interface they have been working with, rather than the user being sent to another state site with a different interface and navigation mechanism.

Another application was developed to aid relief efforts following the September 11th terrorist attacks. It includes syndication to the U.S. Postal Service Web site to retrieve address information when a zip code is entered. Using syndication, zip codes were electronically converted into city and state addresses, reducing staff time for data entry.

There are many Web sites that provide valuable information and services that can be incorporated into other applications. Some examples include weather, lottery scores, news feeds, and more.

In addition to using XML for moving data between Web sites, XHTML is also being used to prepare for extended future use of XML. XHTML in a new W3C standard (http://www.w3.org/TR/xhtml1/#xhtml) that blends XML and HTML. While XHTML allows compatibility with current browsers that support HTML, it begins to build compatibility for future browsers that will support XML.

Technical Details. The New York State portal application is written in Macromedia's ColdFusion. In the example of the local officials, a template was placed on the source Web server, which is activated when there is a request from the portal site. When activated, the template executes a query against a database. The query results are packaged using WDDX, which is an XML wrapper protocol, and sent to the portal site where the data is unpackaged and reformatted, based on the requirements of the application. In the other two examples provided, information is scraped from the source site, packaged again using WDDX, then parsed and reformatted as appropriate.

While these methodologies all provide information to the user, the preferred method to use is the first example, where a template is placed on the source Web server. This approach enables the most accurate, up-to-date information to be culled and promotes increased communication between the "host" and the "source."

The net result of using syndication and XML is that the user never knows that the content did not come from the portal site they are currently using. This simplifies and improves the user experience.

New York State is researching the possibilities of using additional syndicated information as it moves forward with the Governor's "Government Without Walls" initiative. This e-commerce initiative was conceived as a way to enhance access to government services, by providing access at anytime, from anyplace,. This need was brought into sharp focus during and after the World Trade Center (WTC) disaster. WTC illustrated that e-commerce

gives government the ability to continue to provide services and transactions when citizens cannot use traditional methods, such as visiting a government office. The state will continue to develop services for its businesses and citizens that allow them to get the information they need, when they need it.

For more information, contact Barry Meehan, NYS Office for Technology, Applications Team by telephone at (518) 402-2017 or via e-mail at barry.meehan@oft.state.ny.us

National Institute of Justice's XML Solution for Information Sharing

By Dr. John D. Hoyt
Program Executive
Information Technology
Office of Science and Technology
National Institute of Justice

The National Institute of Justice's (NIJ's) InfoTech program has made significant advancements in applying XML technology to support interoperability in law enforcement. InfoTech is an information-sharing technology initiative that has sponsored IT development and implementation in law enforcement since 1996. Early project vision took into consideration both operational needs and emerging state-of-the-art technical capabilities. In an effort to provide timely and correct information to law enforcement officers, the InfoTech program set goals to increase situational awareness by providing context-driven information access. Further goals were to increase officer safety by highlighting critical information, to increase effectiveness by providing information quickly regardless of officer location, and to provide access to existing law enforcement sources through rich information search, retrieval, and display technologies. The InfoTech team did not plan to provide just another single-source repository requiring maintenance and additional operational costs. NIJ was committed to providing a low-cost, secure solution that would allow agencies to keep their existing systems, while still providing rich cross-agency information search and retrieval capabilities.

The primary InfoTech developed data-sharing solution allows participating jurisdictions to seamlessly and securely share critical information. Authorized users in involved regions can perform single point queries against multiple, dissimilar data sources. In San Diego, California, for example, the sources consulted with a single query include the Officer Notification System, County Wants and Warrants, National Crime Information Center (NCIC), the California Department of Motor Vehicles, and a regional booking photograph source.

Based on emerging criminal justice community interest in XML solutions, the InfoTech system was modified to allow XML inputs to the central query-mapping component. This capability is being evaluated in San Diego, where an ongoing project is developing the interface between the San Diego Police Department's Computer Aided Dispatch (CAD) system and the underlying InfoTech system. This capability will enable dispatchers to conduct a single query against many valuable legacy sources through their current CAD user interface; currently dispatchers must log in to each data source separately to perform a query. These legacy data sources are a mixture of relational and network model DBMS's.

In order to provide third party application access via an XML interface, InfoTech technology accepts a socket-based XML interface. For this type of installation, the external application uses communication sockets to send XML-coded requests to InfoTech and to accept XML-coded responses. This protocol supports user interfaces such as an interactive user interface: This might be a computer-aided dispatch (CAD) system, an investigator's query system, or other system through which a user issues a query. The "external application" (from InfoTech's view) is initially responsible for accepting information from the user and formatting it into a legitimate InfoTech query, sent to InfoTech's data fusion module (DFM) via a new-query request. The external application formats the summary return and displays the result to the user. Subsequent drill-down or refresh actions on the part of the user cause the external application to send the appropriate request to DFM and to format the results appropriately.

In addition to providing an interface capable of supporting XML queries, the InfoTech team has developed a rich data dictionary of law enforcement standard data objects that can be used to generate an XML DTD valuable to many criminal justice agencies. The InfoTech team has found that it is possible to define a set of grassroots data objects that are applicable to state and local criminal justice agencies across the United States. The same data dictionary has been used to share regional data in California, Oregon, Virginia, Florida, and South Carolina, requiring few changes to bring a new agency online.

Even in testing stages, InfoTech technology generated leads and information that improved officer safety and enhanced the abilities of law enforcement professionals to bring criminals to justice. In a significant vote of confidence, several regions requested follow-on efforts that they are internally funding. The InfoTech team was also nominated

for a Community Watchdog Award in California, as a project responsible for saving taxpayers money (based on personnel time savings and a cost-effective technical approach).

For more information, contact Dr. John D. Hoyt, Program Executive - Information Technology, National Institute of Justice, Office of Science and Technology by telephone at (202) 616-1471 or via e-mail at HoytJ@ojp.usdoj.gov.

Wisconsin Justice Goes XML!

By Dean Ziemke
Strategic Account Manager
Department of Electronic Government (DEG)
Bureau of Justice Information Systems (BJIS)
State of Wisconsin

The Wisconsin justice community has embraced XML (eXtensible Mark-up Language) as an important data transport vehicle. The Wisconsin Consolidated Court Automation Programs (CCAP), formerly the Circuit Court Automation Program, has several interfaces in production that use XML, and they have participated on the national Legal XML groups in establishing court electronic filing standards.

CCAP uses XML to transfer data from 69 county circuit court databases to the Court Information Repository (CIR) in a central location. They also transfer delinquent income tax warrants from the Department of Revenue to the 69 county databases.

The three largest counties (Milwaukee, Dane, and Waukesha) use XML DTD's to transfer court schedule information to the prosecutor's offices. Two of these counties file an electronic version of their criminal complaint with the clerk of circuit courts office. (Paper filings are still required by law, but data entry is reduced in both offices.)

The statewide prosecutor case management system, PROTECT, developed by the new Wisconsin Department of Electronic Government's Bureau of Justice Information Systems (BJIS) has just begun a pilot implementation of this same electronic filing of criminal complaints. The circuit courts, through CCAP's software, return the court case number, court event information, charge amendments after filing, and charge disposition. Court calendars and case dispositions and sentencing information will be included in the next phase.

Criminal court case dispositions are already being transmitted to the Crime History Repository in the Wisconsin Department of Justice Crime Information Bureau using XML. The charge dispositions are linked with arrest charges submitted earlier from booking agencies. This XML transfer has been operational for over two years. CCAP and the Department of Corrections are also planning the XML transfer of case sentencing information reflecting an electronic version of the Judgment of Conviction.

The Crime Information Bureau has also developed an Internet-based search system called eTIME. eTIME currently searches the Department of Transportation records for drivers' licenses and vehicle registration records. XML data is returned from the query and displayed in HTML, using XSLT to translate the XML data. This data would also be available to be processed and stored by the receiving agency. Criminal history reports are the next phase of this project.

The Department of Electronic Government's BJIS is also developing a Law Enforcement Interface to transfer XML data from local law enforcement and county jails to the county district attorney's offices.

The Wisconsin Justice Information Sharing (WIJIS) project is examining the use of Simple Object Access Protocol (SOAP) as an envelope around the XML responses as a statewide justice data transport standard. The existing CCAP XML interfaces would be "SOAPified" and the other interface would be designed using SOAP for either publish/subscribe or query/response.

As further data transfers become a reality, XML and SOAP will continue to be an important way to provide consistency with future adaptability.

For more information, contact Dean Ziemke, Strategic Account Manager, Department of Electronic Government (DEG), Bureau of Justice Information Systems (BJIS), State of Wisconsin, by telephone at (608) 261-2437 or via e-mail Dean.Ziemke@deg.state.wi.us.

XML in the Wisconsin Courts

By Jean Bousquet
Chief Information Officer
Wisconsin Court System

The Wisconsin Consolidated Court Automation Programs (CCAP), formerly the Circuit Court Automation Program, was an early adopter of XML technologies. Early in 1999, CCAP developed an HTTP-based XML transport facility dubbed Simple Transaction Exchange Protocol (STEP). CCAP's first STEP-based XML interface was implemented in 1999 for the purpose of electronically transferring criminal disposition data from the Wisconsin circuit courts to the DOJ's criminal history database. Detailed STEP documentation can be found at http://stepmom.courts.state.wi.us/overview/.

CCAP has fully embraced XML and related technologies and has incorporated them into many aspects of the court system's application architecture. Currently implemented solutions using XML technology include several inter-agency data interfaces, production of all circuit court forms and reports, and court database synchronization. A SOAP-based query/response mechanism is in development that can be used by Web-based software applications to access CCAP's database of statewide public court information. In the planning stages is the use of STEP and XML in support of an electronic court filing system.

Inter-agency Data Interfaces

Using the custom-developed STEP transport technology, CCAP has developed and implemented several publish/subscribe data interfaces with Wisconsin state agencies.

- DOJ Criminal Disposition Interface. This interface supports the electronic transfer of criminal case disposition information from the Wisconsin circuit courts to the Department of Justice for the purpose of populating the criminal history database.
- DOR Tax Warrant Interface. This two-way interface is initiated when the Department of Revenue publishes tax warrants that are retrieved by CCAP and routed to the appropriate circuit court for the purpose of creating a tax warrant case. When the case is disposed, satisfaction information is sent to DOR, and the DOR database is updated.
- District Attorney Interface. This interface is currently in production in Wisconsin's three largest circuit courts: Milwaukee, Dane, and Waukesha. Criminal cases are initiated in the courts via information provided by the DA's in the form of an XML message. The CCAP applications automatically publish XML case information whenever there is activity on a case (calendar events, docket events or disposition information), which is retrieved by the DA case management system in order to update their databases with current case information. This interface is also in pilot with the statewide prosecutor system in Iowa County.
- DOT Traffic Citation Interface. Currently in pilot in two Wisconsin counties, this interface is used for several purposes. First, State Patrol officers, using laptop computers in their squad cars, create an XML citation for every citation issued. At the end of the day, the officers go to any circuit courthouse and use a custom application to publish these XML citations. The STEP system routes them to the appropriate circuit court. Another application is used by the clerk of court to view these citations online and create traffic cases on the CCAP case management system. As various events occur on the traffic case (such as disposition), XML messages are sent to the State Patrol and the DOT in order to update their databases with current information.

Currently in the planning stages are interfaces with the Department of Workforce Development, the Department of Corrections, and the Public Defenders' Office.

Court Document Production

The CCAP software development team created a document generation subsystem that is used to create, display, and print court documents using XML, XSL-T, Xpath, and XSL-FO technology. This subsystem has been successfully implemented in CCAP's case management system (CMS) in 36 counties with plans to transition the remainder of the counties by June 2002.

- XSL-T is an XML-based language for the purpose of transforming an XML document into another text-based document.
- Xpath is an expression language used by XSL-T to access or refer to parts of an XML document.
- XSL-FO is an XML-based vocabulary for the purpose of specifying detailed page layouts.

CCAP generated documents fall into two general categories, forms and reports. Forms are those documents that have very specific formatting requirements. Examples of such documents in Wisconsin are: Judgment of Conviction, Juror Qualification Questionnaire, various court notices, and Order for Financial Disclosure. There are dozens of these types of forms, whose format is specifically defined by a records management committee.

Reports are those documents that calculate, summarize, tabulate, or otherwise group information into some useful format. Examples of such reports are: General Ledger, Jury Panel Member List, Juror Voucher Report, and Divorce Disposition Report. As with the forms group, there are dozens of such reports available for printing from the CMS.

CCAP's CMS is a suite of Java-based three-tier applications (not to be confused with a Java Applet). Each circuit court is provided with a dedicated relational database server. The applications communicate with the database server through a custom-developed middle tier using SQL statements packaged as Java objects.

CCAP began designing its new printing engine in late 1999 using the following simple criteria:

- The system needed to be operating system independent; this requirement effectively ruled out anything from Microsoft.

- If possible, the system needed to use non-proprietary document formats.
- If possible, the system needed to use open source software.
- The system needed to support the separation of document content from document presentation.

As it turned out, a printing engine was developed that perfectly met all desired criteria. This engine was based on Apache's Xerces XML parser, Xalan XSLT processor, and the FOP formatting object processor. All of these tools are high quality, free of charge, and readily available with access to the source code.

Database Synchronization

CCAP centrally maintains a large database called the Central Information Repository (CIR). CIR is essentially a conglomeration of data as it exists in each independent circuit court database. The CIR database is the warehouse of data that is accessed by CCAP's public access Web site, Wisconsin Circuit Court Access, (WCCA). It is also used for many other internal functions, including providing statewide statistical information.

XML and STEP are used as the vehicles to keep CIR in-sync with changes to the circuit court databases. As changes (updates, inserts, deletes) are made by court personnel in a county, XML messages that represent the changes are packaged up and sent to the STEP server for processing. A separate application removes these XML messages from STEP and applies the corresponding database transaction to CIR. Currently, STEP processes approximately 200,000 database transaction XML messages per day.

SOAP-Based Query/Response

CCAP's Wisconsin Circuit Court Access (WCCA) offers the public access to detailed court case information through its freely available Web site http://ccap.courts.state.wi.us. This site is extremely popular and currently receives over 200,000 requests per day.

Many of these requests come from companies that use custom software applications to access the site for the purpose of interfacing with their own databases or applications.

Since HTML is primarily a presentational format, it is extremely difficult to extract relevant information from an HTML (standard Web page) document.

In response to this problem, CCAP is actively developing a parallel means to access the public case information using the XML-based SOAP-standard. When complete, companies (and the public) will be able to issue SOAP requests to WCCA using the same search criteria currently available through existing HTML forms. SOAP responses will be sent in the form of XML documents containing the same result information currently provided through HTML.

By using the power of XML to provide well-formed data to the public, the content (data) will be separated from the presentation information, making it much easier to programmatically access public court information through WCCA.

Electronic Court Filing

CCAP is currently designing an electronic filing system that will provide a mechanism for attorneys and/or other litigants to file documents with the court without ever creating a paper copy or physically entering the courthouse.

The CCAP system will "read" information directly from an XML-based electronic document and, after verification by the clerk, appropriately update the case management database. As currently planned, only the content of these court documents will be stored—no digital representation of the physical document will be maintained. This methodology will be utilized for several reasons, including:

- Document format obsolescence. Storing only data ensures that documents can be re-created in a "future" document format. Storing physical representations of documents invites future problems when those formats become obsolete.
- Additional search capabilities. Storing document data opens up vast searching possibilities. Groups of documents can be searched, sorted, and otherwise organized using any information contained in the document.
- Physical storage requirements. Much less disk space will be required to store such files, since no presentational information will be stored.

CCAP, as an early adopter of XML technologies, is continuing to eagerly embrace and support the use of XML-related technologies in the court system and government in general. The benefits of using XML will increase proportionately with the number of organizations that implement XML solutions. CCAP can confirm, through much experimentation and production implementations, that the benefits of implementing XML technologies are many.

For more information, contact Jean Bousquet, CIO, Wisconsin Court System, State of Wisconsin, via e-mail at Jean.Bousquet@courts.state.wi.us.

XML in a Citizen-Centered e-Government: *Having Our Cake and Eating It Too*

by Owen Ambur
Co-Chair, XML Working Group
Federal CIO Council

If something sounds too good to be true, it usually is. There's no such thing as a "free lunch," and raising false expectations is a sure path to disappointment, if not outright failure. So the last thing we should do is oversell the benefits of eXtensible Markup Language (XML). As has been often said, XML is merely syntax—a way of structuring information. Moreover, structure without substance is meaningless—or worse.

In many instances, ***too much*** structure is more a problem than a solution. We've all experienced hierarchies that have outlived their useful lives and cannot cope with rapidly changing realities in the cyberage, much less add real value in the best interests of their stakeholders. However, as inefficient and ineffective as inflexible, poorly designed, and outmoded structures can be, some degree of commonality is required for the conveyance of meaning, not only among human beings, but especially among machines. It is a truism that we cannot manage our own lives, much less interact productively together in organizations large and small, without some degree of structure.

And that generalized truism becomes an absolute requirement in the realm of machines. It has been widely suggested, for example, that the beauty of the Web is its "openness" and lack of structure, particularly the lack of any centralized authority controlling what may be posted. However, without standards like Internet Protocol (IP), Hypertext Transfer Protocol (HTTP), and Hypertext Markup Language (HTML), the Web just ain't happening. The truth is that the beauty of the Internet and the Web is in their structures. For it is those standards that enable us to use computers to express ourselves more openly and efficiently, thereby freeing the collective creativity of the human spirit.

Apart from gestures that can be interpreted logically and nonverbal expressions that seem to be universal regardless of culture, human beings simply cannot communicate without a common language in which words and phrases have widely recognized meanings. XML is not a "language" in that sense, since in and of itself it does not convey meaning. However, it enables us to "mark up," that is to add structure to our documents, not only to make them more readily comprehensible by people, but also to enable computers to analyze, manipulate, display, re-purpose, and reuse documents in ways that are highly useful to us. And that is what technology should be about—serving the needs and wishes of people, thereby enabling us to lead "freer," more productive, meaningful, prosperous, enjoyable, and fulfilling lives.

As human beings, we have a marvelous and, indeed, a wondrous ability that computers lack. We can fill in gaps and apply our own understandings and interpretations where the full express meanings are lacking. Of course, to the degree that our perceptions and interpretation may differ from that which was intended by others, those capabilities can get us into a lot of trouble, including not only hard feelings and physical conflict on a personal level, but also law suits among business partners and, ultimately, war among nations. In effect, we need to help ourselves avoid such unintended consequences—by helping those poor "dumb" computers do a better job of helping us understand each other.

So what does all of this mean to us in our daily lives as government employees? What should we be doing about it? And how can we work most effectively together to capitalize on the potential of XML in pursuit of our agency missions and priorities established for us, such as those outlined in the Administration's Citizen-Centered E-Government Action Plan?

Two of the original proposals that led the CIO Council to charter the XML Working Group included the potential for: 1) using XML metadata tags to classify, manage, access,

Index to Federal News

and retrieve electronic records, including Web pages, government-wide, and 2) rendering government forms in XML and gathering the data from those forms in XML.

Rendering forms in XML means that they can be filled in with a Web "browser." Gathering the data in XML means that it can readily be captured, manipulated, and analyzed in databases. At the same time the original, completed forms can be maintained as inviolate records for the appropriate periods specified in agency records retention schedules. Maintaining the original, completed forms as e-records apart from the databases provides for redundancy in the event of disaster, a requirement made more gravely apparent by the events of September 11th. It also enhances security and the protection of privacy in cases where personal or other sensitive data may be required for authentication of each submission, but may not be appropriate for inclusion in a database where it is subject to the risks of inappropriate access, modification, and use.

Equally important, maintaining records apart from the databases facilitates audits of those databases. Indeed, if the original e-records are gathered and maintained in well-formed XML, audits can be largely automated. This means that auditors can devote their attention and talents to much higher-value activities than tracking down pieces of paper to confirm numbers in spreadsheets. It also means that any stakeholder will be able to conduct his or her own "audit" of the public records anytime he or she sees fit to do so! And that's a big part of what "citizen-centered" e-government is all about—providing accountability by making the actions of public officials transparent to their stakeholders.

An equally big part of citizen-centered government is making our records readily accessible by our stakeholders—in terms that are meaningful to them, rather than merely to us as employees of various offices within the bureaucracy. Pursuant to Section 508 of the Rehabilitation Act, as amended by the Workforce Investment Act, the term "accessibility" has come to mean enabling the rendition of information in a manner that is comprehensible to persons whose sensory organs are diminished or disabled. That is certainly an important benefit of XML, which enables the ready reformatting and reuse of electronic records via specialized devices such as computer screen readers. However, it assumes people are able to discover and retrieve the information they need without great difficulty, regardless of whether they are disabled or not. And that is where the use of XML metatags can deliver large gains in efficiency and effectiveness for all of us, by enhancing search and retrieval time as well as precision.

Metatags are the equivalent of the information provided in a library card catalogue. They identify the attributes of each record that are of significant interest to its stakeholders. By embedding metatags in records on the Internet, agencies can automate the management and processing of their records, while at the same time vastly enhancing the services provided by existing, full-text search sites like FirstGov.

Since there is no limit on the number of metatags that may be included in each record, agencies can indeed "have their cake and eat it too." Agencies can incorporate not only those metadata elements required to describe their records in terms that are most meaningful to the members of the public who are their stakeholders, but also in terms that are important for internal administrative purposes. That is a critical distinction relating to the use of metatags. Metatags can be embedded within individual records and maintained by "suppliers" on a widely distributed basis. At the same time, they can be dynamically and selectively used in other contexts by "customers" (citizens), as well as by third-party organizations serving more specialized needs and interests.

By contrast, static indices of hypertext links are limited by screen space, reflect the context of the supplier rather than the customer, and are difficult to keep current on a centralized basis. However, XML does not force an either/or choice between metatags that are embedded within widely distributed records versus external indices maintained by librarians, authoritative organizations, and other value-added service providers on a centralized basis. By virtue of its "extensibility," XML enables the efficient and cooperative use of both embedded and external metadata.

Indeed, externalized XML metadata may be used to construct "topic maps" that identify relationships and draw linkages among records that may not have occurred to and/or are beyond the capabilities or interests of their authors to specify. For example, an agency may not have the resources to identify many of the relationships of its records to those created and maintained by other agencies. However, commercial and nonprofit organizations representing those agencies' stakeholders may be more than willing to do so. Rendering agency records in well-formed XML enables others to provide value-additive indexing services efficiently and effectively, building upon, without having to re-create, the work already done by the agencies themselves. That's another example of how XML can enable us to "have our cake and eat

it too" with respect to overcoming the false choice between embedded versus external indices.

While XML is still a relatively new technology and the tools to apply it efficiently are immature, it is incumbent upon each of us as public servants to be thinking about and planning how best to use XML to address the needs and interests of citizens. Toward that end, the XML Working Group has been chartered by the CIO Council to undertake four general activities:

- Identify pertinent standards and best practices
- Foster partnerships with key industry and public interest groups
- Foster partnerships among existing communities of interest within government
- Engage in results-oriented education and outreach

The Working Group meets monthly. Our meetings are open and participation is encouraged. Likewise, our listserv is open, and we are hosting the xml.gov site as a primary means of sharing information with our stakeholders. However, beyond generalized awareness-building, education, and outreach activities, the primary value the Working Group can add is to accelerate the availability of a repository in which "inherently governmental" XML data elements and schemas can be registered and made readily available for use in applications government-wide. The registry is the key to reducing, if not eliminating, needless inconsistencies and redundancies among the data stored in ill-coordinated information technology systems that do not inter-operate and, thus, constitute "stovepipes" or "islands of information." Moreover, eliminating needless inconsistencies and redundancies is the key not only to making government more efficient and effective, but also to reducing the information collection burdens imposed upon the public.

Unfortunately, life is not as simple as it used to be, and the ever-increasing complexities seem at times overwhelming. No technology can insulate us completely from the new realities we face. However, XML embodies a powerful new potential to represent both what we all have in common, as well as what makes each of our organizations and us unique. Unlike many resources, knowledge is not "consumed" in the sense of being used up or converted to waste. The more knowledge is used, the more valuable it becomes. While there is no such thing as a free lunch, if we play our XML schemas right, we ***can*** "have our cake and eat it too," in terms of an evermore rapidly expanding base of common and specialized knowledge, leading inexorably to a brighter future for all of us—suppliers and customers, public servants and citizens alike!

For more information, contact Owen Ambur via e-mail at owen_ambur@fws.gov.

EBS Web Services

By Drew Anderson
U.S. Army

The Army is using eXtensible Markup Language (XML) in many ways. The Electronic Bid Solicitation (EBS) provides a good example.

- 1. The Central Reporting System uses XML as a message format. Local Web sites communicate with our **Tri-Service Solicitation Network** (tsn.wes.army.mil) site, informing us of updates to their sites.
- 2. Mailing labels are created using a Word macro that hits our Web site to figure out where all the local Web sites are. Then once a user has selected a local Web site, a list of advertised solicitations is downloaded. The user selects a solicitation, and then the Word macro creates the mailing labels using data sent from the Web site. All communication between the Word macro and the Web sites is done with XML.
- 3. Web Services is allowing industry to poll our local EBS Web servers for solicitation, in much the same way that we poll the Web sites to post their solicitations to the Army Single Face to Industry (ASFI) Acquisition Business Web Site (http://acquisition.army.mil/default.htm).
- 4. XML is also used by the U.S. Army Engineer Research & Development Center to assist Districts in debugging their EBS Web sites. We have even written a program that allows us to match their database structure with the one we have here.

Striving to simplify the transfer of Electronic Solicitations to industry, EBS Web sites have been outfitted with Web services. The Web services expose an Application Program Interface (API) that publishes solicitation information. The API works like a function normally found in a program. A request is made specifying the method to execute and pass the necessary parameters. A response is then formulated, based on the request, and sent back. The whole process is performed using the http protocol over the Internet. Requests to the API are formatted in XML, as are the corresponding responses.

Following is the pseudo code for obtaining all solicitations, including files, from EBS Web sites.

- Get a list of EBS Web sites.

- For each site:
  - Get a list of Advertised Solicitations.
  - For each Solicitation:
    - Get Solicitation Meta Data.
    - Get a list of Amendments.
    - Get a list of Plan Holders.
    - Get a list of Solicitation Files.
  - Next Solicitation
- Next Site

Utilizing Microsoft's Document Object Model (DOM), applications can be easily written to process this information. To use Microsoft's DOM in a Visual Basic project, one must first add the 'Reference': Microsoft XML, Version 2.0, or higher. Following is an example of how to obtain a list of all EBS Web sites with Web services.

```
1. Dim doc As New DOMDocument
2. Dim node As IXMLDOMNode
3. Dim attr As IXMLDOMAttribute
4. doc.async = False
5. doc.Load "http://tsn.wes.army.mil/EBSWebSites.asp"
6. For Each node In doc.documentElement.childNodes
7.  For Each attr In node.Attributes
8.  Select Case UCase(attr.nodeName)
9.    Case "NAME"
10.  Debug.Print "Web Site Name: " & attr.nodeValue
11.  Case "URL"
12.  Debug.Print "URL:" & attr.nodeValue
13.  End Select
14.  Next
15.  Next
```

Line 1 creates a new DOM object. When loaded with XML data, it represents a tree of data.

Line 2 defines a 'node' object. A 'node' could have 'child nodes,' 'attributes,' and/or data. In this case, we are only concerned with the 'child nodes' of the root node.

Line 3 defines an 'attribute.' Here we only use the 'nodeName' and 'nodeValue' properties of an attribute.

Line 4 tells the DOM object to download the entire XML data stream before processing it.

Line 5 does the actual http call to the given URL. The XML data the script produces is loaded into the DOM and made available for us to read. At this time, the Microsoft DOM object does not support SSL.

Line 6 will loop through the 'child nodes' of the root node. Since the structure of this XML data stream is only two levels deep, this is all that is needed to process all of the data.

```
<Response .....>
   ; Root node with attributes
   <WebSite ........\>;
      Child nodes with attributes
   <WebSite ........\>
   <WebSite ........\>
</Response>
```

Line 7 will loop through all attributes of a node. A node could have many attributes, each having a name and a value.

```
<WebSite              ; Node name
   Name="..."         ; Name of EBS Web Site
   URL="..."          ; URL to Web Services
   Database="..."     ; Database name in use at the site
   Link="..."         ; URL to Solicitation Web page
   Code="..."         ; Solicitation procurement code
   FilePath="..."     ; Path to solicitation files/>
```

Lines 8 through 13 will print each Web site's name and URL.

As you can see, using Web services is really very easy. One just needs to know how to make the call to the API and then understand the format of the data coming back. For a complete listing of available API function calls, please refer to http://tsn.wes.army.mil/TSNSoftware.asp.

For more information, contact Drew Anderson by telephone at (601) 634-4226 or via e-mail at Drew.L.Anderson@erdc.usace.army.mil.

A Brief History of SGML and XML at the United States Patent and Trademark Office

By Bruce Cox
U.S. Patent and Trademark Office

The United States Patent and Trademark Office (USPTO) currently publishes approximately 3,500 patent grants per week and nearly that many patent applications. Sometime in 2002, the number of applications is likely to exceed the number of grants published. Long before the volume of applications reached its current level of more than 300,000, the United States, Japan, and the European Patent Offices (Trilateral

Offices) realized that they all faced the same difficulties—a rapidly growing workload that could not be supported by simply enlarging the workforce. Automation and its ability to improve efficiencies were high on the list of possible solutions. In 1993, USPTO tasked a contractor to advise the USPTO on the use of SGML in processing patents. Aside from the initial education of USPTO staff on the finer points of SGML, the principal accomplishment was the completion of the first draft of the World Intellectual Property Organization (WIPO) Standard ST.32, which included the DTD for published patent documents as agreed among the Trilateral Offices.

Starting in 1995, a matrix-style "SGML Transition Team" continued educating USPTO staff and planning for the eventual use of SGML. The Transition Team developed a Concept of Operations, an Implementation Guide, and a Validation Mechanism. The Implementation Guide included draft SGML DTDs developed for U.S. patent applications and other filing-related documents. In 1996, a unique DoD- and USPTO-funded DARPA project offered the opportunity to create a substantial test bed for many of the SGML-related ideas, including the DTDs for SGML-tagged patent applications. The Distributed Object Computing Testbed (DOCT) project demonstrated that it was feasible to use SGML throughout the prosecution of an electronic patent application. Coincidentally, the process converted the back file of photocomposed published patents (1975-1998) from a flat text format to SGML.

With the emergence of several SGML-dependent projects in the USPTO, the Chief Information Officer revived the Transition Team in the summer of 1998 to harmonize the use of SGML throughout the organization. Shortly thereafter, a new project began to develop a system that would make extensive use of XML to process patent applications after filing.

A matrix team was formed in 2000 to coordinate the development of the USPTO's Electronic Filing System (EFS) DTDs and the DTD for publishing patent applications 18 months after filing. Applications are created using a specially prepared word processor that enforces the constraints of the DTD. Since April 1999, the USPTO has been publishing patent grants using an SGML DTD based on WIPO Standard ST.32. SGML is also used to publish the Manual of Patent Examining Procedure (about 2,000 pages of rules for patent examiners and applicants), as well as other publications used by examiners and patent applicants. ST.32 and USPTO patent grant publications will migrate to XML in 2002.

In 2001, the Trilateral Offices and the World Intellectual Property Organization initiated international coordination of XML resources for electronic filing of electronic patent applications. The new international DTDs and filing protocols were published for final review in November. (The USPTO will implement these new DTDs in October 2002.) Based on that success, the Trilateral Offices agreed to coordinate XML resources for prosecution of electronic applications and all applicant-office communications. This unprecedented level of international cooperation will significantly reduce the effort required of applicants who submit applications to many industrial property offices. It also sets the stage for reducing workload at the Trilateral Offices by making it technically possible for Trilateral Offices to exchange work products.

The discovery of anthrax-contaminated mail in autumn 2001 continues to seriously delay delivery of mail to and from the USPTO, which in turn has significantly spurred interest in electronic filing. Currently, the USPTO is seeking vendors to create authoring tools for applicants. The tools will produce documents that conform to the new international DTDs and then facilitate filing using the agreed upon protocols, much as the IRS did for electronic filing of tax returns.

For more information, contact Bruce Cox, Director, Information Products Division Chair, XML Technical Working Group via e-mail at bruce.cox@uspto.gov. USPTO XML resources and documentation are available at http://www.uspto.gov/ (select Site Index, then R, and then click on Red Book). The XML resources jointly developed by the Trilateral Offices and WIPO are available at http://pcteasy.wipo.int/efiling_standards/EFPage.htm.

AEC/XML DATA Model for Equipment Manufacturers

by David H. Horner, Civil Engineer CADD/GIS Technology Center for Infrastructure, Facilities and Environment

The Facility Maintenance and Operations Committee (FMOC) of the National Institute of Building Sciences (NIBS) is collaborating with the U.S. Navy, the National Institutes of Heath (NIH), and the CADD/GIS Technology Center to promote and develop an eXtensible Markup Language (XML)-based data format standard for equipment manufacturers. The final goal is to provide a uniform data standard for defining building equipment that can

be directly integrated with any Computerized Maintenance Management System (CMMS).

XML is the universal format for structured documents and data published on the Web. XML is a set of rules, guidelines, and conventions for designing text formats, in a way that produces Web-enabled files that are easy to generate and read. XML formatted text files are unambiguous and avoid such common pitfalls as the lack of extensibility, inability to support internationalization/localization, and the limitation of platform-dependant applications. Hypertext Media Language or HTML was the first format language used on the Web. With added functionality, it progressed to Standard Generalized Markup Language (SGML). The designers of XML simply took the best parts of SGML and produced something that is no less powerful than SGML, but is vastly more regular and simpler to use. However, while SGML is mostly used for technical documentation and much less for other kinds of data, XML is exactly the opposite. A data structure diagram describing the place where XML-SGML-HTM fits and how it is structured is shown below.

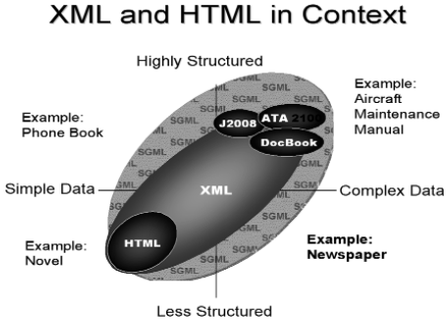


Figure 1. XML-SGML-HTML based on the structure of the data

Project Development: The FMOC and its partners have developed a feasible and robust model that is supported throughout industry and by the federal government. Information submitted by manufacturers has been successfully authored in and/or converted to AEC/XML, and reasonable cost estimates for authoring documents and converting existing documents have been obtained. Software developers such as Archibus and Bricsnet have demonstrated capabilities to read and apply the data within their applications.

In recognition of the Navy's efforts to fully implement electronic documentation of building information, the project incorporates the Navy's Operation, Maintenance, and Support Information (OMSI) data requirements. Alliances have also been forged with the Machinery Information Management Open Systems Alliance (MIMOSA) and the International Alliance for Interoperability (IAI).

This effort will be coordinated with other NIBS-supported efforts including the U.S. National CADD Standard and Total Building Commissioning. A successful project will encourage proper attention to the larger issue of standardization of data across the entire life cycle of facilities. In many instances the challenge is simply to exercise the forethought to capture needed data already created for prior needs in the facility life cycle.

The primary database and structure (schema) can be found at ftp://tsc.wes.army.mil/FMOC.

Demonstration Process: Demonstration project(s) will be selected to identify the steps required to fully implement the data standard in a CMMS. Candidate buildings will be selected from the agencies that have participated in development of the data model. Project funding will be used to apply the model. The architecture-engineering and/or construction contractor will also be funded to provide project information for the data model. Information will first be provided in SGML and later in XML. NIBS will contract with firms experienced in SGML to apply the submitted project data to the data model. The vendors of the CMMS, used for the demonstration project buildings, will develop data exchange interfaces to allow use of the project data within the context and functionality of their software products.

The vendor-developed interfaces demonstrate the process of importing and exporting XML-defined data in the OMSI into and out of CMMS maintenance applications. This can be seen in the following diagram and three-sided pyramid showing the interaction between information and outcome.

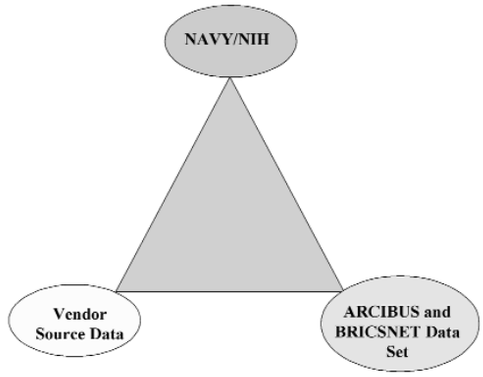


Figure 2. Interaction of data and outcome

We are also working with the IAI's AEC/XML initiative, which is where the NIBS' and the Navy's efforts fit within the IAI structure. The FMOC is moving towards product information data applications. This is very important for equipment manufacturers, who often provide field engineering to complete the installation of their systems. The result is often product information too general for use by building managers and technicians, who need in-depth information. The

following diagrams represent a detailed plan of what has been accomplished and what is going to be done on the project. The Navy started with OMSI and that has been integrated into a complete and usable electronic format. It is this format that will be integrated with XML for use by engineers, developers and managers for viewing and retrieving information. That information will shorten the daily requirements for Facility Managers.

For more information, contact David H. Horner by telephone at (601) 634-3106 or via e-mail at hornerd3@wes.army.mil .

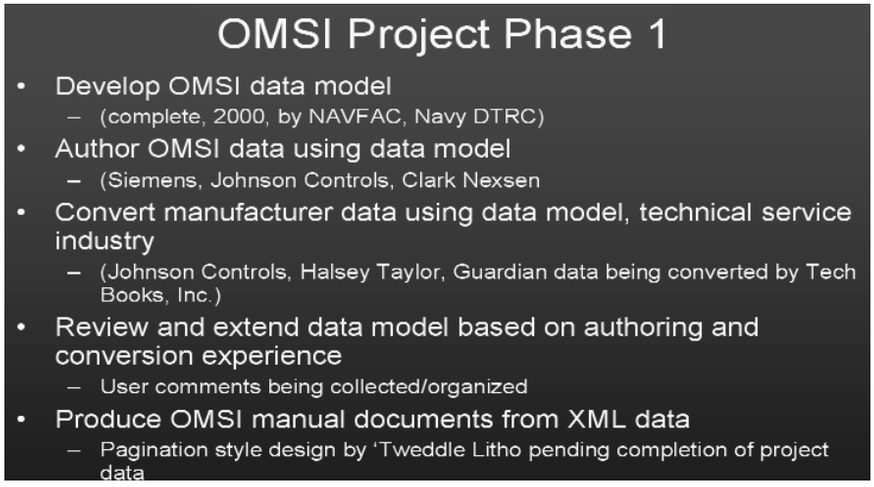


Figure 3. Phase 1 of OMSI Project

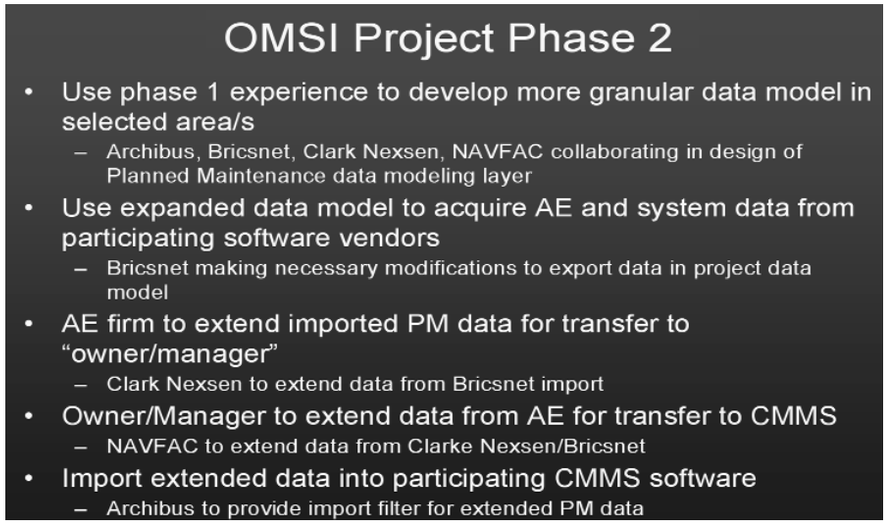


Figure 4. Phase 2 of OMSI Project

Department of the Navy XML Work Group

By Michael Jacobs
Department of the Navy

As is true throughout industry and the federal government, the Department of the Navy[1] (DON) is striving to efficiently and effectively implement XML across its enterprise to improve application integration, eGov connectivity, content/knowledge management, and collaborative process execution. Leading this effort for the Department is the DON XML Work Group. It is charged with:

> **...providing the leadership and guidance to maximize the value and effectiveness of emerging XML component technologies implemented across the DON enterprise.**

In order to achieve this mission, the Work Group plans to develop a comprehensive set of XML "discovery documents," policies, guidance, and training materials. The first two discovery documents, an *Interim Policy on the Use of XML for Data Exchange* and an *XML Developer's Guide,* have been completed and released. The next document to be released will be an *XML Primer.* The purpose of these documents is as follows:

- I*nterim Policy on the Use of XML for Data Exchange:* The *Interim Policy* directs the use of World Wide Web Consortium (W3C) official recommendations and directs DON XML developers to register all new XML components that they use. This policy was signed-out on September 6, 2001.
- *XML Developer's Guide:* The *Developer's Guide* provides DON XML developers with implemen

tation direction, such as naming conventions and style guidance. The guide is grounded in standards and specifications from ebXML and ISO11179. The initial version of the *Developer's Guide* was signed out on November 1, 2001.

- *XML Primer:* The purpose of the *Primer* is to provide a high-level overview of XML component technologies, as well as to highlight the potential enterprise benefits of XML implementation. The *Primer* has been drafted and is undergoing review.

A significant challenge facing the Work Group is ensuring that the selected technical approach to XML implementation will yield the greatest improvements in business interoperability over the long term. The Work Group feels that a **joint** business/technology XML vision is crucial to achieving this outcome. To that end, the Work Group will encourage active and ongoing collaboration among business and technical stakeholders in the definition, development, and delivery of XML policy and guidance to the DON enterprise.

The Work Group is also aligning its efforts with other federal, industry, and consortium initiatives, such as ebXML, OASIS, OAGIS, and the Federal CIO Council XML Working Group.

The DON XML Work Group is comprised of approximately seventy participants from across the Department of the Navy and industry, including a number of XML consortium members. Additional information about the DON XML Work Group can be found at http://quickplace.hq.navy.mil/navyxml.

For more information, contact Michael Jacobs by telephone at (703) 601-3594 or via e-mail at jacobs.michael@hq.navy.mil.

1 The Department of the Navy includes both the Navy and the Marine Corps.

Program Executive Office (PEO) Interchange XML Initiative (PIXIT)

By Dr. Joseph J. Molitoris
Chair, PEO Interchange XML Initiative

What is the PEO Interchange XML Initiative (PIXIT)? It is a working group within the PEO Interchange. What is the PEO Interchange? This MITRE-enabled effort began with meetings between the Army and Air Force C2 (Command and Control) Program Executive Office (PEO) in late 1999. Navy and DISA participation was added around May 2000, and the Marine Corps joined the fray in December 2000. The current executive participants are Army PEO C3S, Navy SPAWARSYSCOM, USAF PEO for C2 Programs, DISA Information Engineering, and USMC Systems Command. The PIXIT research described here was performed over the past year and presented to the PEOs and MITRE groups on several occasions.

What is XML? XML, as a method of structuring data in a text or Unicode character file, forms the potential common link between many disparate and stove-piped systems that are now deployed by DOD, FAA, IRS, and other government agencies. Since XML is text, it can be edited by an ordinary text editor such as Microsoft Notepad or any UNIX text editor [1].

Although now a "techie" thing, XML is rather broad in its scope and promise as an emerging technology. After considerable debate, the PEO Interchange XML Initiative (PIXIT) has concluded that the PEOs would benefit from focusing on XML as an enabling technology to achieve better interoperability [2]. This paves the way for the pervasive use of XML within flagship programs over the next few years, as well as the planned 2002 release of XML-MTF (message text format [3]). Our conclusions and reasoning may apply to other sponsors that MITRE supports, so we provide the business and technical case for XML use here.

As Major General Steven Boutelle of PEO C3S recently noted in an interview with Military IT magazine: "The Web browser is here to stay. XML allows us to take the next giant step in managing information."

XML is one of several web technologies that are emerging now, some of which can be leveraged by the Army, Air Force, Navy, and within other government agencies. The PEOs are just beginning to feel the force of directives, policy, and mandates from the Pentagon that relate to these Web technologies. The reasoning behind this initiative is that what works effectively for so many millions of users should help improve military interoperability as well. Furthermore, XML is a family of technologies consisting of the basic XML specifications of 1998, as well as more recent World Wide Web Consortium (W3C) and other recommendations on Schema, StyleSheets (CSS and XSL), Namespaces, transformations (XSLT), and other matters [4].

Is XML ready for prime time? The basic XML specification of 1998 came from working group efforts of 1996. The second edition of that specification was published in 2000. There are a number of related recommendations that were released by W3C over the 1996-2000 period. Major software companies

(Microsoft, Sun, IBM, etc.) have implemented the specification in their systems. Only in the encryption/security and query area are the specifications not fully mature, but that development is expected sometime in 2001. Thus, XML is ready for prime time with respect to basic XML functionality, style sheets for viewing, document object model for programming access, wireless applications markup and protocols, and DTDs/schema for data/information specification.

What is the recent history of XML from the point of view of a MITREITE and refugee scientist? In the mid-1990's, the Internet and World Wide Web became the hot technology for everyone, including military and government personnel. One of the enablers there was HTML (hypertext markup language). Because of the simplicity of HTML—it is not a programming language, you only need a text editor to write it, and you only need a browser to view the display—the Web moved forward exponentially from a few million users to hundreds of millions of users today.

XML, like HTML, is rather simple. XML is not a programming language, although it may evolve many of those capabilities. One needs only a text editor to write it and a browser to view the display. Sure, the techies will extend (read "improve") it and add more complicated twists as they did with HTML (this is already happening), but this need not take away from the simplicity and utility of XML. XML is an extended markup language which means that the programmer, developer, or a community of parties define common tags (technically elements) that enable better interoperability and easier information exchange [5].

When should XML be used? XML should be used by government programs and agencies where it is mandated and/or where it makes sense. XML is not a silver bullet for all information processing woes, but it has the potential to increase interoperability if used intelligently. Currently, XML is mandated for tagged data items by the JTA (joint technical architecture) and JTA-A (JTA-Army) within DOD. There are at least one hundred people working on XML projects throughout MITRE and even more efforts within other companies and numerous government agencies [6].

XML should be used in flagship programs, messaging, databases, office products, and other processing where information transfer between disparate systems is important

and where interoperability requires improvement. XML need not be used in all cases, but should be considered as a document/data definition format in the same way that HTML is considered and utilized as a display format. Furthermore, XML can also help with the display of data with the new XHTML specification and the transformation properties of CSS, XSL, and XSLT.

What is the potential impact of using XML? The proper use of XML can save the government money in system deployment and increased interoperability. The potential benefits include cost savings, reduced training time, interoperability within each service (or agency), and interoperability between the services (or agencies).

The use of XML can occur at several levels:

- Internal to a particular system
- Between two or more single service/department systems
- Between two or more multiple service/department systems.

The last two categories could be called the external use of XML, and coordination between the affected programs/communities of interest is important. This coordination can occur informally or formally and is discussed later under the topic of communities of interest (COI). The current commercial models for coordination (consortia, partnerships, and standardization groups) are all good models for defense and government departments.

Do you have to rewrite entire software systems to utilize XML? No. XML may be incorporated into the appropriate pieces of a software system's architecture in the same way that HTML and browsers were incorporated in the 1990's. Naturally, if one were just beginning a software system or light (PC) version of an existing system, then it would be easier to implement XML than with an existing system that was not designed for Web technology. And the depth of penetration of XML into the architecture may relate to the amount of interoperability achieved.

However, flat files and database queries can be converted to XML and vice versa. Agencies that have an installed based of SGML should have an easier time with transitioning to XML than those with only HTML experience.

Can One buy XML? No. One can buy many products that support XML (and some are freely available), including browsers, operating systems, office products, databases, and portals. But someone has to produce the XML that a particular organization needs, unless the specific XML already exists in a commercial or government repository.

Would one have a hard time using XML? **Not necessarily. The basics of XML are fairly simple. The hard part is the human agreement between the developers or managers of disparate systems who need to agree on the tag or element names. Thus,**

within DOD, communities of interest (COI) need to form agreements to use XML. These COIs parallel the consortia, standards bodies, and corporate partnerships that occur in the commercial marketplace to forge agreements about the common use of XML. A COI can be as small as a single program within a PEO or as large as several PEOs working together with respect to a particular mission area. If agreement is not possible, then it is still easy and fairly cost effective to design translators between different XML systems.

When should XML not be used? When the system or group of systems already operates acceptably and XML use is not mandated or considered appropriate.

What does one need to get started using XML? As with any Web or software development project, a clear vision of requirements and data/information exchange needs is helpful. In terms of tools, the following items at a minimum are needed:

1. Text editor (e.g. Microsoft Notepad on Windows or a UNIX text editor on a UNIX system)
2. An up-to-date browser (Internet Explorer, Netscape Navigator, or others with XML support)
3. Parser, unless the parsing in the browser is sufficient (e.g. XML Spy, MSXML, Larval)
4. DTD or schema editor unless the DTD/schema checking of the browser is sufficient (e.g. XML Authority)
5. (optional) XML enabled database tool for database work (from Oracle or other vendors)
6. (optional) XML programming language access (XML Document Object Model or other interface)

In using XML, what text editor or development environment should be used? Microsoft Notepad or any plain text editor will work. The user creates XML files in accordance with the XML specification. HTML or SGML experience is particularly useful for a new XML developer. There are also more expensive development environments from several vendors (Extensibility, SoftQuad, and i4i, to name a few). Some of these, such as the i4i product, are plug-ins to Microsoft Word for ease of use.

In using XML, which browser should be used? Different browsers have different levels of support for XML. It is fairly safe to use the latest version of Internet Explorer (version greater than 5.0) or Netscape Navigator (version 6.0 or above). Most modern browsers have at least limited support for XML.

In using XMLs, should schemas or DTDs be used? This is the choice of the developer. Many products already use DTDs, but it is not hard to convert to schema should they be needed. Schema are in general more capable and more complicated, but light versions of schema such as the RELAX (REgular LAnguage description for XML) approach are available as well. A recent space ontology schema [7] has been developed for the Air Force by MITRE, and an evaluation of the different schema products has also been completed [8]. Extensibility makes an XML DTD and schema editor called XML Authority that is fairly easy to use and relatively inexpensive.

What areas of XML activity are appropriate for DOD and government agencies? Our recommendations to the PEOs have potential application throughout DOD (and, indeed, in other government departments and systems). In order for XML (a COTS technology) to be successfully adopted (technology transition) for military use, four areas must be addressed:

1. Policy - DOD, DISA, Army, Air Force, Navy, Marine Corps, NATO or affected departments and agencies and agencies
2. Cross-service interoperability
3. Programmatic, e.g. PMs within Army, Navy, Air Force, Marine Corps
4. Technical – Actual technical implementation.

Success with XML (unless it occurs by grassroots activities, as it did with HTML) must address these four problem levels.

From Stovepiped to Interoperable Systems. We have shown elsewhere [2] how military systems are making use of browsers and Web markup languages such as HTML and XML. More universal browser use as a common interface and the wider use of HTML and XML should be planned for. This will help complete the movement from stovepiped and proprietary to Internet standard based systems, thereby enabling greater ease of use and interoperability.

The defense and government departments should plan to more efficiently use available limited bandwidth as systems are fielded. As faster computers and greater bandwidth become available, these should be deployed when and where cost effective in order to ease the strain on processors and networks.

As lessons are learned from experimentation with Web technologies like XML, the government should move from these demonstrations and proofs of concept to fielding of the best and most cost effective products. This is now in progress with several programs.

Interoperability, the Holy Grail. Interoperability is defined in different ways by different agencies. Here, two definitions within the joint military community from Joint Publication 1-02 are noted. Of course, interoperability is important for companies and the consumer/user/warrior as well.

At the simplest level, interoperability means 1) providing and accepting services (data, messages, information) and 2) exchanging information directly and satisfactorily between two or more systems.

The Internet and Web technology [5] have allowed us to achieve considerable interoperability over the last decade. XML allows us to move a step further. XML is aimed most obviously, but not exclusively, at the data level of interoperability [9]. Our challenge is to achieve better interoperability between government systems for our sponsors. Interoperability could encompass multiple message formats, file formats, data and database formats, and operators. XML is designed for this type of heterogeneous environment.

A Multiyear Vision. XML is already being used in prototype efforts in every community of interest, from the FAA [10] to the IRS and DOD [11]. Within DOD, efforts range from those within the intelligence community [12] to messaging standards bodies [3].

Following is a vision of XML's place in PEO (and other agency) systems beyond the next few months. The application of XML to PEO systems over the next few years has great potential to improve information exchange and interoperability. First, it can work with message processing (CMP and IRIS). Then it can work with database exchange, beginning with mission critical databases. Finally, it can work with a potentially more integrated joint C3 architecture.

Major software companies—such as Microsoft, Sun, IBM, Sybase, Informix, and Oracle—are integrating XML within their products in the same way that HTML was integrated just a few years ago. This is because XML is a powerful method of structuring data in a text or Unicode character file. XML files can handle data/text information, as well as pointers to binary files. As such, XML can be important for information exchange between applications and/or between clients and servers.

Following are recommendations for three areas where XML should be applied (where appropriate):

- Flagship programs desiring to better synchronize and interoperate. This is a two-part recommendation: 1) messaging and 2) databases. Messaging is the easier task in XML, enabling the CMP and IRIS products.
- Leveraging COTS XML software (browsers and parsers, operating system and databases, other tools.
- Developing, supporting, and implementing XML policy and guidelines.

It is very important that the policy efforts at the top (DOD and service CIO groups) are consistent with the intelligent application of XML by developers to improve software products. This will only occur with good feedback and communication between all stakeholders. The PEO Interchange effort and the working groups can help make this happen.

Acknowledgements. I want to thank and acknowledge contributions from Frank Ruscil, MITRE Eatontown and Ed Masek and Dan Winkowski of MITRE Langley. At the executive level, I am grateful to Beth Meinert, Danny Demarinis, MITRE Eatontown, and Dr. Joe DeRosa, MITRE Bedford.

References

[1] XML Elements of Style, Simon St. Laurent, NY: McGraw Hill, 2000.

[2] *XML Technology*, Dr. Joe Molitoris, MITRE Eatontown, November 8, 2000, and other more recent briefings and publications including the MITRE IT Adviser, March 2001.

[3] *XML-MTF*, Mike Cokus, MITRE Langley, several briefings and documents.

[4] See www.w3.org for further information.

[5] Web Technology briefings by Bruce Binney (MITRE San Diego), Lynne Spruill, and Dr. Joe Molitoris expand on this area.

[6] A listing of known XML efforts is available on request.

[7] Space Ontology Schema, Dan Brandsma and Mary Pulvermacher, MITRE Colorado Springs.

[8] Relational Databases and XML, Alan Bernstein, MITRE Eatontown, January 2001.

[9] Mike Bean, MITRE Bedford.

[10] Jon Mack, MITRE Atlantic City, work in progress.

[11] *XML Within Government*, Dr. Joe Molitoris, MITRE Eatontown, December 2000.

[12] Tim Croy, MITRE Maryland, January 2001, work in progress.

Acronyms

BFA	Battlefield functional area
CMP	Common message processor
COTS	Commercial off the

	shelf
CSS	Cascading style sheets
DISA	Defense information systems agency
DOD	Department of defense
DTD	Document type definition
HTML	Hypertext markup language
JTA	Joint technical architecture
JTA-A	JTA-Army
PEO	Program executive office
PM	Program manager
RELAX	Regular language description for XML
SGML	Standard generalized markup language
XHTML	Extensible HTML
XML	Extensible markup language
XSL	Extensible stylesheet language
XSLT	Extensible stylesheet language transforma tions
W3C	Worldwide web consortium

For more information, contact Dr. Joseph J. Molitoris by telephone at (732) 389 6705 or via e-mail at jmolitor@mitre.org.

Building Peer-to-Peer XML Content Networks of Web Services for Federal Scientific and Statistical Data and Information: *FedStats.Net and Beyond*

By Brand L. Niemann
U.S. Environmental Protection Agency(EPA)

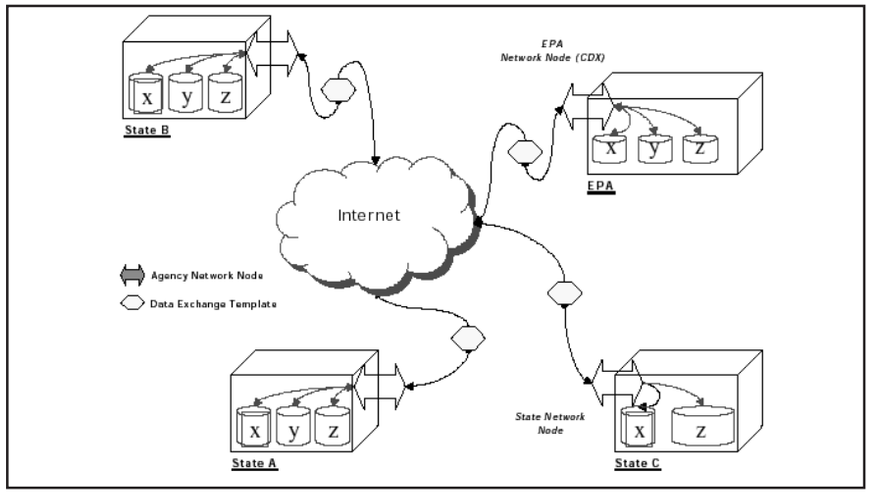
XML is a new standard for preserving and communicating information—encoding, tagging, and internationalizing on the World Wide Web. Web services use XML data and messages to provide communication between applications running on different Web servers that will bring the Internet to its new level (called the third generation). Peer-to-Peer (P2P) architectures are best understood by thinking in terms of Presence, Identity, and Edge (PIE) resources that offer powerful approaches for solving the seemingly intractable problems of the Web and for creating a more efficient and robust Internet.

P2P presents the most significant challenges to traditional centralized IT departments and security models. XML data and metadata architectures are connected because (1) from a "resource perspective," metadata is either external to or embedded within the data, while (2) from a "client perspective," metadata is either centralized or distributed. Peer-to-peer approaches are needed in a more distributed environment to simultaneously query, discover, and aggregate the data and metadata results locally.

Peer-to-Peer XML Web services is the technology underlying new programs like the EPA National Environmental Exchange Network and Central Data Exchange, content networks, and several e-government initiatives. Everyone in the Federal government should know something about Peer-to-Peer XML Web services.

Introduction

The EPA goal is to create a National Environmental Information Exchange Network (see figure below), an Internet-based system where standardized environmental data is rationally shared and easily



accessible. With EPA and each state having one point of entry, or node, and following agreed-upon formats, EPA and the states see many benefits. These include reduced data redundancy, better-quality data, improved security, and, ultimately, better environmental policy [1].

The National Environmental Exchange Network has been cited recently as an excellent example of "Achieving the Vision of e-Government" [2], because from a technology perspective, messaging and Web services have become the key enabler to convergence.

1. Global trade requires global communications and increasingly involves e-markets, e-procurement, online product development, supply chains/value Webs.
2. Trade requires data exchange and information processing capabilities from data storage, to transaction processing, to standard protocols.
3. Legacy and new systems are coded in multiple languages (COBOL still dominant, C++, Java, Ada, etc.) and were modernized by Y2K remediation.

So, XML.ORG and the World Wide Web Consortium (W3C) have embraced messaging standards (SOAP) and "services" (UDDI) to allow communications between standard platforms and a broad range of legacy applications.

This means that users must have a unified view of data and simple business processes in order to reap benefits from Emerging Web Services Technologies. Further, e-government opportunities create benefits throughout the Information Value Chain: Capture, Store, Query, Distribute, Analyze, Act, and Learn [2]. Specifically, these new technologies and opportunities:

1. Reduce burden—for example, applying online for services, submitting data and payments online.
2. Reuse information, consolidate databases, and unify islands of automation to reduce maintenance and usage costs—e.g. peer-to-peer or search vs. data warehouses.
3. Publish information to devolve governance to local communities.
4. Improve knowledge management to improve decisions, yielding better service, faster and at lower costs.

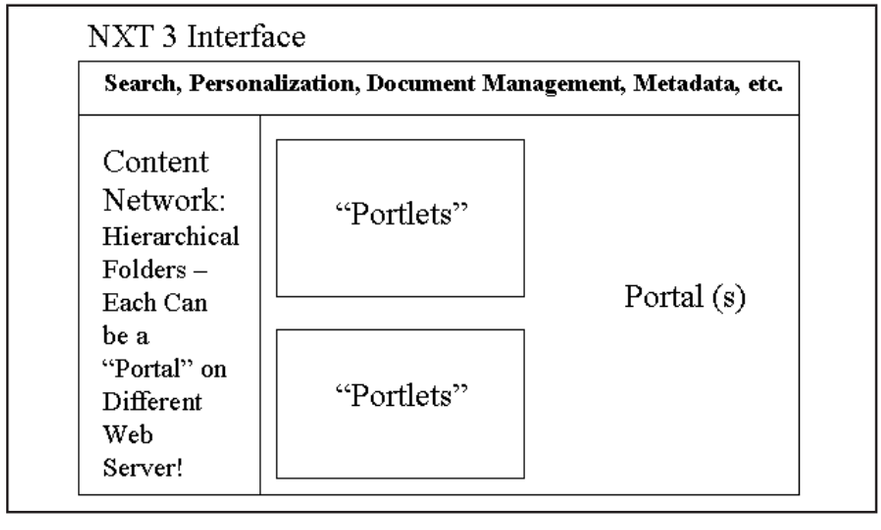
This paper reports on some pilot efforts to build Peer-to-Peer XML Content Networks of Web Services for Federal Scientific and Statistical Data and Information. It begins with an effort called "An Integrated Repository and Registry for Environmental Information" [4], then FedStats.Net [5-9], an outgrowth of the FedStats.Gov activity [10-16] and more recently called the Pilot FedGov Content Network [17-23].

Virtual Centralization of Diverse and Distributed Content

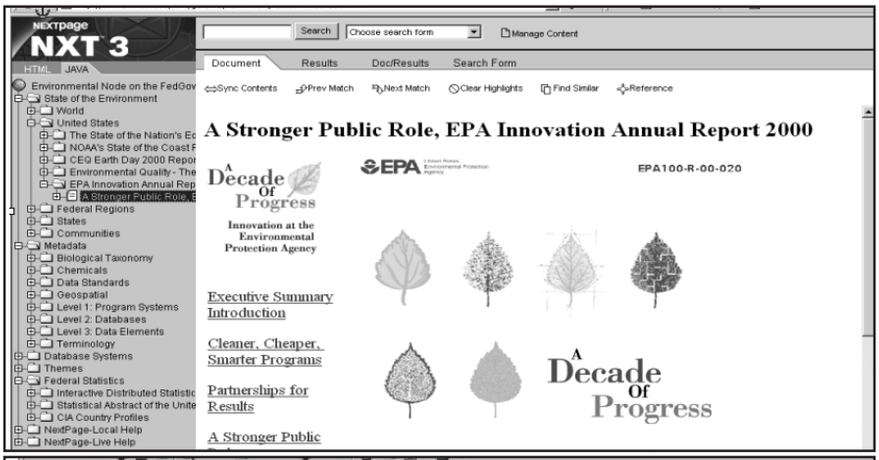
A recent Gartner Group Report [24] describes the emergence of distributed content management and Peer-to-Peer Content Networks. The Integrated Repository and Registry for Environmental Information [4] and FedStats.Net 5-9] used NXT 3 P2P Platform from NextPage to pilot the use of messaging between Web servers and Web services that integrate information and applications for environmental and Federal statistical data and information. The NXT 3 interface (see Figure below) combines contents networks (left pane), portals and "portlets" (center pane), and services like search, personalization, document management, and metadata (top banner). In a business application, the "portlets" are usually e-mail, stock, and news feeds. In our environmental and statistical applications, the "portlets" are passing the XML data to mapping and statistical analyses applications. The content network system of hierarchical folders (left pane) can be likened to the Windows Explorer tool where diverse file types can be organized into folders. However, with NXT 3, the diverse file types and folders can and should be on different Web servers where the content originates and is maintained.

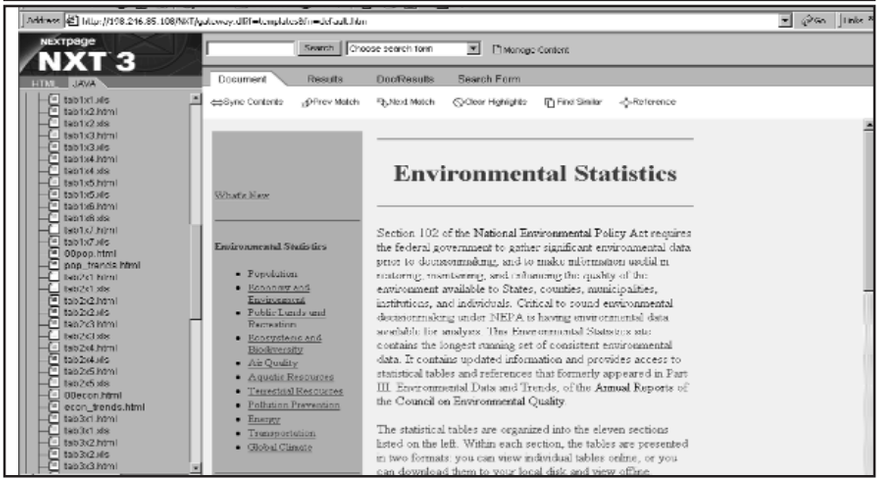
Environmental Node on a FedGov Content Network

The Integrated Repository and Registry for Environmental Information [4] (see Figure below) has grown to become the Environmental Node on



the FedGov Content Network. This includes the FedStats.Net [5-9] content, plus other content added more recently to demonstrate the versatility of the NXT 3 P2P Platform. The environmental information is organized by what users want to know, namely, what is the state of the environment, how good is the data the state of the environment is based on (called metadata for data about data), what are the databases used in the assessing the state of the environment, and what is the best information about the environment organized by themes.





Permanent Public Access Archive

The NXT 3 P2P Platform has also been used to pilot (1) the archiving of distributed and diverse content to provide "permanent public access" and (2) the creation of new content nodes with added-value through the use of XML and Web services. This has been done with selected environmental and Federal statistical and scientific content like the Environmental Statistics (see Figure below), which archives about 200 HTML and Excel files that are updated periodically. Section 102 of the National Environmental Policy Act (NEPA) requires the Federal government to gather significant environmental data prior to decision-making and to make information useful in restoring, maintaining, and enhancing the quality of the environment available to states, counties, municipalities, institutions, and individuals. Critical to sound environmental decision-making under NEPA is having environmental data available for analysis. This Environmental Statistics site contains the longest running set of consistent environmental data. It contains updated information and provides access to statistical tables and references that formerly appeared in "Part III -Environmental Data and Trends" of the Annual Reports for the Council on Environmental Quality.

Pilot FedGov Content Network

The pilot FedGov content network was assembled by using the NXT 3 P2P platform to link multiple Web servers with distributed content. The purpose was to test the messaging functionality of making distributed content look and function as though it was all on one Web server, when in actuality it was on multiple Web servers (see Figure above). At one point in the pilot work, we had diverse content on up to six Web servers, including the extremes of Web servers in California and Washington, D.C., and servers next to one another.

Value Proposition for the Federal Government

The value proposition is:

1. XML "future proofs" information against periodic technology change, facilitates integration, and promotes collaboration.
2. P2P reduces costs of integration and can replace data replication and warehousing.
3. Web services provide communication between applications running on different Web servers that will bring the Internet to its new level.

The NXT 3 P2P Platform has provided excellent support for our

XML/P2P/Web Services pilot projects.

XML Web Services Training

A system of XML Web Services training has been developed for both managers and specialists to support the EPA Central Data Exchange and Region and State Network Node building process. The multiple levels are:

1. Overview
2. Level 1 – Introduction to XML Video
3. Level 2 – Introduction to XML for the Web: Step-by-Step Hands-on Class
4. Level 3 – Basic Training in the EPA Network Node Pilot Projects: Alpha Phase
5. Level 4 – Specialized XML Asset Management for CDX and the Network: Overview of XML Spy 4.0 and TIBCO Canon/Developer/Portal
6. Level 5 – XML Web Services Security Models and Technologies for Use in CDX and the Network
7. Level 6: XML-ization of Databases and Embedding and/or Linking to Data Standards, Metadata, and Data Quality Information
8. Level 7: Building XML Web Services at the national, regional, and state levels

References

1. A Networked Environment, Federal Computer Week, Greg Langlois, June 18, 2001, http://www.fcw.com/fcw/articles/2001/0618/pol-epa-06-18-01.asp
2. Achieving the Vision of E-Government, Mark Forman, Associate Director for IT and E-Government, OMB, October 2001, 18 slides.
3. The XML Files and XML Project Centralizes Agency Statistics, Federal Computer Week Article, January 8, 2001, http://www.fcw.com/fcw/articles/2001/0108/cov-xmlbx3-01-08-01.asp
4. Niemann, B. 2000: An Integrated Repository and Registry for Environmental Information: An XML Portal for Public Access and Data Exchange, Discussion paper for INFOTERRA 2000-Global Conference on Access to Environmental Information, Dublin, Ireland, 11-15 September. http://www.unep.org/infoterra/infoterra2000/Brand-rev.doc
5. Niemann, B.L., FedStats.Net, November 2000, Federal Conference on Statistical Methodology,
6. Niemann, B.L., et al, FedWeb 2001 Conference, Portals: How E-Government Is Transforming Communications With Citizens: From Portals to Peer Space, and The Future of Portals: Case Study of FedStats as a Model for Collaboration and Data Integration, March 2001, http://www.fedweb.org/FedWeb2001Presentations/Future%2520of%2520Portals-FedStats%5b1%5d.ppt and http://www.fedweb.org/FedWeb2001Presentations/Portals%2520to%2520Peer-to-Peer%5b1%5d.ppt.
7. "Peer-to-Peer Isn't Dead", The Industry Standard Magazine, April 23, 2001
8. "NextPage Bridges Islands of Content", ComputerWorld, May 28, 2001, http://www.computerworld.com/cwi/story/0,1199,NAV47-71-365-383_STO60858,00.html
9. "All in a day's work" and "Computers with a view", Federal Computer Week, http://fcw.com/fcw/articles/2001/0514/mgt-profile-05-14-01.asp and, http://fcw.com/fcw/articles/2001/0514/cov-fednap-05-14-01.asp.
10. Capps, C., Green, A. and Wallace, M., 1999: The Vision of Integrated Access to Statistics: the Data Web, Of Significance, Volume 1, Number 2, 42-47, Association of Public Data Users.
11. DeBerry, M., Gregg, V., and Taylor, R. FedStats: Creating the U.S. National Statistical Information Infrastructure of the 21st Century, UN/ECE Seminar on Integrated Statistical Information Systems (ISIS), Riga, Latvia, 29-31 May 2000.
12. Marchionini, G. and Hert, C., 2000: Interacting with Tabular Data Through the WWW, in proceeding of this conference (http://istweb.syr.edu/~tables/).
13. Wallace, M. and Cortez, J.A., 2000: Developing and Promoting Next-Generation Inormation Products for WWW.CENSUS.GOV and WWW.FEDSTATS.GOV, Census Advisory Committee of Professional Associations, October 19-20, 12 pp.
14. Wallace, M. and Highsmith, S. 1999: Use of Metadata for the Effective Integration of Data from Multiple Sources, ICES paper, 8pp.
15. Wallace, M., and Sperling, J. User-Driven Integrated Statistical Solutions (Version 08/28/00), accepted for publication in URISA Journal, http://www.urisa.org/Journal/accepted/sperling/user_drivewn_integrated_statistical_solutions.htm
16. Wallman, K.K., Zawitz, M.W., Blessing, C., and Treadwell, W., 1999: Making Things Add Up For the End Uses: Issues in Statistical Literacy, Of Significance, Volume 1, Number 2, 14-16,

Association of Public Data Users.

17. Niemann, B.L., XML/P2P/Web Services in EPA, XML.GOV, & E-GOV, O'Reilly Peer-to-Peer and Web Services Conference, November 5-8, 2001, 10 slides.
18. Niemann, B.L., Building P2P/XML Content Networks of Web Services for Federal Scientific and Statistical Data and Information: FedStats.Net and Beyond, 2001 S-PLUS International User Conference, October 18-19, 2001, 41 slides.
19. Niemann, B.L., 2001 Federal Depository Library Conference, Agency Update: USEPA, October 15, 2001, 33 slides.
20. Niemann, B.L., Maximizing Peer-to-Peer Portal Through XML: An Integration Case Study from EPA, Enterprise Web and Corporate Portal Conference, September 5-6, 2001, 49 slides.
21. Niemann, B.L., What is XML?: A practical introduction for EPA Webmasters, EPA Web Working Group Meeting, July 24-26, 2001, 40 slides.
22. Niemann, B.L., XML@EPA, EPA National Developers Conference, April 3-5, and Veterans Administrations, April 10, 2001, 31 slides.
23. Niemann, B.L., XML for Information and Application Integration, Federal CIO XML Working Group Presentation, http://xml.gov/fedstats_for_xmlgov/index.htm.
24. The Emergence of Distributed Content Management and Peer-to-Peer Content Networks, http://www.nextpage.com/products/tours/530b_final.html, http://www.nextpage.com/news/pressroom/pr_030701.html, http://www.nextpage.com/gartner%20whitepaper%20final.pdf

For more information, contact Brand L. Niemann, Ph.D by telephone at (202) 260-2510 or via e-mail at niemann.brand@epa.gov or http://www.sdi.gov/server.htm

The OJP XML Reconciliation Effort, or A Tale of Three Specifications

By Donna Lindquist
XML Working Group
IIR, Inc

Introduction

The Office of Justice Programs (OJP), U.S. Department of Justice, eXtensible Markup Language (XML) Reconciliation Effort reads like a good story. At the heart of the tale is a dilemma, a threatened warring of wills. And, like all good endings, potential conflicts were diffused, people came together, and everyone wound up better off than when they started. But it started with a problem. And that problem was sharing information.

The Problem

For public servants striving to protect our nation's citizens—whether at the local level or homeland security scale—the ability to exchange critical data throughout the entire justice community is no longer a "wish list" item. It is a necessity. The one-two punch of modern transportation and advanced technology has given rise to geographically boundless crime. Furthermore, entities outside of the justice community that impact our well-being—schools, child care services, transportation, and licensing agencies—also need this critical information to perform daily business activities, such as hiring new personnel, approving gun purchases, or granting professional licenses.

The Response: The XML Reconciliation Effort

In response, a number of efforts are underway to develop XML-based data exchange specifications within the justice and public safety communities. However, OJP observed that rather than working toward a mutually agreed-upon solution, the efforts were independent. In fact, in the worst cases, they were counterproductive.

To address these issues and foster interoperability among emerging specifications, OJP, in cooperation with the Infrastructure/Standards Working Group (ISWG) of the Global Justice Information Network (Global), identified three emerging specifications to engage in the XML Reconciliation Effort project.

The specifications selected were the Interstate Criminal History Transmission Specification, developed by the Joint Task Force on Rap Sheet Standardization; the Regional Information Sharing Systems (RISS) XML Data Exchange Specification, developed by RISS; and the Electronic Court Filing Proposed Standard, developed by LegalXML. (Brief backgrounds on these specifications follow in the Notes section.) Additionally, standards specialists from the Institute for Telecommunication Sciences, National Telecommunications and Information Administration (NTIA), played important roles as technical advisors to the process.

The reconciliation effort officially kicked off in March 2001, with an XML Technology Working Group (Working Group) meeting in Kansas City, Missouri. Early work centered on analyzing the three XML specifications. The result: It was obvious that three excellent efforts had been

undertaken—independently. It was also clear that each practitioner group aimed at satisfying its own mission and operational requirements. Therefore, the XML implementations specified for the Rap Sheet, RISS, and the Electronic Court Filing Standard were incompatible as originally envisioned. For example, tagging conventions were different, as well as the data elements themselves and how they were used.

Through a series of meetings of the XML Working Group and XML Focus Group—a subcommittee of selected representatives from the projects—principles and procedures were established and agreed to by all participants. From the outset, it was recognized that the needs of the individual practitioner groups were of the utmost importance. However, during the technical discussion, it turned out that one practitioner group would have strong feelings about a particular technical approach in one functional area, but not in another, allowing technical compromise and giving the participants the ability to dissect the whole effort and operate on it in "pieces." Also, with three sets of developers looking at information sharing from different perspectives, there was a chance to pick the "best ideas" from each of the specifications, regardless of developer. This willingness to set aside issues of propriety and "turf" was fundamental to the success of the project, especially at such an accelerated rate: Early reconciliation agreements were reached inside of four months.

The Success

Through this spirit of collaboration and persistence, the three XML specifications have been reconciled, enhancing justice information sharing among the courts, local police, and Federal law enforcement. Additionally, two cooperative products have been developed:

1. The ***XML Justice Data Dictionary***, a "living," evolving database, currently containing over 135 reconciled data elements.
2. The ***XML Justice Standards Development Principles***, a set of guidelines and lessons learned from the process. These principles serve both as parameters for the current effort and blueprints for future XML reconciliations. This is important as the project moves forward, exploring the inclusion of other agencies.

These lessons and development principles, as well as a brief history of the participants and reconciliation effort, were combined into the XML Reconciliation Effort ***Lessons Learned*** report. These documents are available on the Global Web site at http://www.it.ojp.gov/global/standards/xml.html.

The Future

Next steps involve submitting the reconciled XML standards and common products to the imminent ISWG Justice Standards Registry Program, a program that will include the cataloging, recognizing, and sharing of pertinent standards, as well as facilitating collaboration in standards development efforts. The registry will be available on the Global Web site, and readers are encouraged to become involved as registered users.

XML Working Group members are also committed to growing the effort into a second phase. This important transition involves expanding the reconciliation of XML standards beyond the three original groups. A logical next partner is the American Association of Motor Vehicles Administrators (AAMVA), considering that the data AAMVA such as information about driver licensing and vehicle registration is certainly relevant to the justice community. Indeed, this collaboration not only benefits involved parties, but also local and state governments and practitioners across America who are pursuing public safety and homeland security by liberating the powerful currency of timely, accurate, and complete information.

As you might expect, this story is to be continued.

For more information, contact Mr. John Terry, Institute for Intergovernmental Research by telephone at (805) 385-0600, extension 291, or via e-mail jterry@iir.com or Mr. D. J. Atkinson, NTIA by telephone at (303) 497-5281 or via e-mail at dj@its.bldrdoc.gov.

Notes: The Players

XML Interstate Criminal History (Rap Sheet) Transmission Specification

In 1996, the Joint Task Force (JTF) on Rap Sheet Standardization—with representation from the Federal Bureau of Investigation (FBI), FBI Criminal Justice Information Services Advisory Policy Board (CJIS APB), National Law Enforcement Telecommunication System (NLETS), the National Center for State Courts, SEARCH, the National Consortium for Justice and Statistics, and state and local law enforcement agencies—was formed to carry forward the work of the National Task Force on Increasing the Utility of the Criminal History Record. The JTF succeeded in developing a standardized criminal history transmission format; a standard presentation format utilizing the XML transmission format; and a concept of operations, which combines criminal histories from multiple sources into a single criminal history.

Regional Information Sharing Systems (RISS) XML Data Exchange

Specification

RISS is composed of six regional centers that share intelligence and coordinate efforts against criminal networks that operate across jurisdictional lines. Each RISS center has from 530 to over 1,300 member agencies, the vast majority of which are at the municipal and county levels.

The RISS Program developed an XML specification to address dissimilar databases and lack of information sharing standards to allow agencies to share their criminal intelligence information with a maximum of flexibility at a reasonable level of development effort and cost. The RISS XML Data Exchange Specification can be easily modified to allow for additional elements and functionality as additional or revised needs are identified.

For more information, visit http://www.iir.com/riss/.

Court Filing XML Specification

The Court Filing XML Specification provides a mechanism by which documents can be electronically filed with a court system. The effort began when the Conference of State Court Administrators/National Association of Court Managers Joint Technology Committee (JTC) formed an E-Filing Standards subcommittee to pursue an Internet electronic filing specification for the courts.

To that end, in December 1999, the JTC voted to partner with LegalXML. This coalition produced the LegalXML Court Filing Standard. This proposed standard was developed to serve as a universal translator among different electronic filing systems. The standard is built on the following tenets:

- This proposed standard deals only with the data to be exchanged, not with electronic filing systems themselves.
- This proposed standard does not attempt to impose common terminology.
- This proposed standard defines the data to be sent with the document to enable a court to identify the document and its sender, and to create the docket entry in the court's case management information system (CMIS) for the filed documents. The standard also includes all data elements needed by a court to initiate a new case in its CMIS, if necessary.

For more information, visit http://www.ncsc.dni.us/NCSC/TIS/JEDDI/Jeddi-1.htm.

For more information on the Institute for Telecommunication Sciences, National Telecommunications and Information Administration (NTIA), visit http://www.its.bldrdoc.gov/Home.html.

To learn more about justice information sharing and OJP Information Technology Initiatives, visit http://www.it.ojp.gov

Extending Digital Dividends: Public Goods and Services That Work for All

.(A summary of a soon-to-be-published CIO Report)

"Information is the currency of democracy."

Thomas Jefferson

The Federal Architecture and Infrastructure Committee of the Federal CIO Council has developed a guide, *Extending Digital Dividends: Public Goods and Services that Work for All,* as a first step in helping public administrators improve program delivery. The guide raises awareness of technological opportunities for spanning the digital divide. It also asks how e-government services today can mitigate the effects of digital disparities by accelerating innovative solutions that can be shared widely at low cost. Three groups are disproportionately affected: Americans with disabilities, older people without work experience using computers, and people with English or technical literacy limitations, compounded at times by both cultural or geographic isolation and limited finances. The guide:

- Identifies those demographic groups that are disproportionately under-served by e-government and, in an effort to cultivate intervention strategies, presents information on the addressable issues that cut across those under-served markets.
- Provides examples of successful e-government efforts to minimize digital disparities.
- Highlights how eXtensible Markup Language (XML) and its variants help Federal agencies improve the delivery of government goods and services.
- Provides background and reference information on technology enhancements, such as voice application networks and networked improvement communities, that are poised to offer forward-looking Federal programs the greatest value as they move to a more citizen-centered e-government.
- Identifies ways for public administrators to implement XML and the other emerging technologies as the Federal government sets in

motion the transformation of a government centered on customer needs.

Communication Tools for a Connected Society. The guide reviews three promising Internet-based technologies: (1) Voice Application Networks, (2) XML, and (3) Networked Improvement Communities. These communication tools have the potential to dramatically extend the reach, quality, and usefulness of public information.

Voice Application Networks. Voice Application Networks (VANs) represent an important new tool for government to use to reach many of the groups currently under-served. Voice Application Networks allow the same dynamic information interactions developed for the Internet to be provided by telephone. In response to a caller's spoken selection, up-to-date information is presented in a natural sounding voice. This new generation of telephone technology allows users to access spoken information via existing telephone lines and equipment using voice-activated applications. The new applications are based on Voice XML and deliver the required information through a combination of programmable interactive voice response and speech recognition. The same information provided on a Web site could be navigated by voice using the telephone, provided that the file structure is XML. Advancements in this technology mean that the voices that callers hear over the telephone are quite natural sounding and easy to interact with. In addition, advances in voice acoustics analysis have allowed providers to build highly sensitive recognition models that are able to account for such things as regional accents and dialects.

An example is Utah's new Traveler Advisory Telephone and Internet system, developed in preparation for the 2002 Olympics. The service will integrate information on weather-related road conditions, accidents, congestion, and construction activities into one system. The same information is available by Web site and an interactive voice recognition telephone system. The dynamic information structure capability of XML makes it possible for current information to be continuously distributed in real-time to both services. Those unfamiliar with the Internet or simply traveling by car can call the 511 abbreviated dialing code. Using a natural sounding voice, the Traveler Advisory service prompts the caller with questions. Callers respond by natural speech to the prompts and are provided the information requested. This approach mitigates the effects of digital disparities by providing the same information through the old familiar channel of the telephone and the newer, less widely available Internet.

Voice Application Networks are maturing in power and popular appeal just in time to be effectively used by localities that are expanding the National "N11" abbreviated dialing services. Today, 411 business directory services and 911 emergency services are most widely used by the general public. Soon more "N11" numbers will become equally known and available. This nationwide dialing system allows telephone users to connect with standard service gateways throughout the country by dialing only three digits.

The City of Hampton, Virginia, opened its 311 call center in September 1999, to provide city residents with 24-hour-a-day, seven-day-a-week access to city services and information, ranging from reporting missed trash collection to answering questions about the city budget. The service has been very well received, and public satisfaction with the service has been extremely high, with the call center receiving an average of 600 calls a day (May 2000).

XML. XML is the starting point for many of the Internet- and voice-based technologies that can be used to improve access to government goods and services and therefore can alleviate many of the barriers to receipt of government information associated with digital disparities. XML itself is a promising notation system for representing the underlying structure and meaning of electronic information, whether it be data, text, or sound. This system and its variants (e.g., Voice XML and Wireless Markup Language) play an important role in providing goods and services directly to the public through telephone-based services with speech recognition or text-based materials on the Internet.

XML functions by employing markers, or tags, that accompany the electronic data in agreed upon ways. These tags, and the structures that govern them, mean that XML can simplify and unify the sources of data to reduce the burden of managing multiple sources of data. These same characteristics allow information to be reused, databases to be consolidated, and disparate databases to be accessed, resulting in lower maintenance and usage costs. Data that is managed once can then be accessed through multiple channels, whether it is shared directly with the public by the Federal government or is shared with a state or local government that then publishes the information.

Many industries and scientific disciplines are already using XML to exchange information across platforms and applications. Because XML separates data from presentation, XML users can extract only the data needed. The implications are enormous for massive manuals,

project schedules, catalogs, etc. on the Web that would be overwhelmingly complex if the specific information needed couldn't be extracted quickly and easily.

Federal agencies that want to take full advantage of XML and its data reuse and re-purposing capabilities must first establish a "vocabulary" for the information to be shared and then organize that vocabulary in an approved manner. Several national and international organizations have evolved to facilitate the efficient use of XML data. These organizations and the standards that have been established ensure that the information managed in an XML format bears the increasingly important characteristics and attributes of a well-formed record: reliability, authenticity, integrity, and usability.

Utah's new Traveler Advisory Telephone and Internet System illustrates the potential of XML. Other states are also using XML. The Michigan State legislature converted its compiled law to XML, which makes it easier to offer the documents online in multiple formats (HTML and PDF formats). This action not only helps state agencies improve the production and management of their legislative process, but also enables the public to access documents without the need for a legal researcher. At some point, the public will be able to see the relationship between legislative actions and the results of those actions expressed in government records. Visitors can draw from different sources to build their own customized output documents. More than 6,000 people use the system every day, with more than 43 million hits since the system was launched two years ago.

The National Archives and Records Administration and the Government Printing Office (GPO) are moving from SGML to XML. XML offers more tools and Web renderings. XML enables applications to: (1) mediate between two disparate databases; (2) present different views of the same data to different users; and (3) support intelligent agents that tailor information discovery processes to the needs of individual users. Agencies that submit documents to GPO receive a 35 percent discount for delivering text that has already been marked up in XML. XML supports the philosophy that data "belongs" to its creators and that distribution channels are best served by a data format that does not bind the content to a particular vendor's tools or platform.

Networked Improvement Communities. Networked Improvement Communities that take advantage of Internet-based collaborative workspaces are one of the best approaches available to government program managers who are looking to create innovative e-government projects that will mitigate digital disparities. Innovative e-government projects exhibit several common characteristics, including the use of public-private partnerships, alliances with stakeholders, interagency cooperation, and a focus on end-users. These characteristics, all of which show openness to accommodating differences, are inherently citizen-centric and fit naturally within the current e-government mandates.

Collaborative workspaces can play a successful role in fostering an end-user focus by providing government program managers with the networked improvement community space needed to build public-private partnerships and stakeholder alliances, as well as to enhance interagency cooperation. Collaborative workspaces do this by combining the functionality of Internet chat rooms with the document storage capabilities of a knowledge management system. These spaces also allow the government to interact with a broader segment of the public, many of whom are currently the most underserved in terms of government goods and services. This interaction can help both entities by allowing underserved groups to become partners in innovation with the government. Together, the partners can identify news tools and technologies that will provide greater access to information.

Collaborative workspaces also allow for the creation of more meaningful partnerships among state and local providers of government services, community groups, businesses, nonprofits, and the public. Businesses are the most sophisticated users of networked improvement communities, which are important tools to manage customer and partner relationships. A good example is the Open Electronic Book Forum (OEBF) that includes traditional media publishers, electronic publishers, hardware manufacturers, software developers, and various other organizations, including disability organizations and access-technology providers. The purpose of OEBF is to create and establish XML standards for electronic book technology and to promote this technology. OEBF books produced by any publisher will be readable on any manufacturer's OEBF-compliant device. The agreements necessary for cross-platform compatibility are critical to success. All partners know that multiple approaches taken independent of one another will result in market failure.

A recent successful government application of the Networked Improvement Community approach was the Environmental Protection Agency's (EPA's) Online Dialogue held in July 2001 that was directed at certain groups, including con-

sumers, environmental and advocacy groups, and minorities. EPA used the online dialogue, which attracted 1,200 users, to gather comments on EPA's draft public involvement policy. The dialogue was structured in a manner similar to an in-person meeting, starting each day in the morning with an online posting of the daily topic, an introduction, and comments made by a panel of experts. Approximately 100 to 200 messages were received each day, subsequently summarized, and then sent to all of the participants. While the dialogue did not replace the traditional Federal Register publication and respective comment period, it did allow EPA another venue in which to gather and share stakeholder comments.

Another successful application of the Networked Improvement Community approach is Earth's 911. Earth's 911 has received more than 70 awards, including the 2001 Stockholm Challenge Award. Earth's 911 is a nationwide public and private sector partnership providing citizen centric environmental information in every community in the nation. Earth's 911 has networked government information, in collaboration with EPA, from 50 states, 3300 counties and thousands of cities. Earth's 911 can be accessed by web at www.earth911.org or bilingual phone at 1-800-Cleanup. Citizens are empowered by XML-based information organized by zip code and available by web or phone.

Summary. The Internet and other information and communication technologies are rapidly changing the way we work, learn, communicate with each other, and conduct government. The most publicly visible metric for e-government success will be enfranchising all who want the opportunity to be involved in local civic life in spite of a handicapping condition such as age, disability, or a previously isolating difference. But the ultimate success of electronic government depends on the ability of all Americans to take full advantage of the opportunities offered by the government. A sound e-government course will tip the market balance toward more commercial offerings available to government and other organizations that are designed for use by all people. As the Internet shifts from presenting static information to dynamic and service-oriented applications, Federal organizations that form active learning partnerships with state and community leaders will tap the power of a citizen-centric approach.

For more information contact Susan Turnbull by telephone at (202) 501-6214 or via e-mail at susan.turnbull@gsa.gov.

International News

The Ministry of Finance of the Czech Republic: Interoperability of Information Systems

By Ji_í Roudn_, IT Director
Ministry of Finance
Czech Republic

The Ministry of Finance of the Czech Republic (MFCR) operates several important information systems integral to the organization's activities. These systems are called "Financial and Control Information Systems" (FCIS) and include the following:

- **Automatic Tax Information System** (ATIS), which administers all kind of taxes that are to be collected in the Czech Republic.
- **Automatic Budget Information System** (ABIS), which monitors how the state budget is being allocated.
- **Customs Information System** (CIS), which monitors the activities of the customs authorities.
- **Central Register of the State Subsidies** (CRSS), which monitors all subsidies granted from the state budget.
- **Program Funding Information System** (PFIS), which monitors the allotment of funds to particular programs of the public sector.
- Other lower priority information systems

In the past, these systems were developed and operated separately, because each was under the supervision of a different division of the MFCR. Another reason was that at the time they were created, namely after the revolution which ended Communism in the Czech Republic and opened the door to new technologies, there was an effort to build the information systems swiftly using new economic and technological resources.

Particular information systems contain basic identification and allocation information on the same objects in the real world, e.g. individuals, enterprises, etc. After some time, however, there was a need for these individual systems to be interconnected, so that one could find all information concerning an entity and display it in one location. For example, one could learn whether the entity has satisfied its tax duties, whether it is receiving a state subsidy, etc. The existence of separate systems created several problems:

- Information on the same entity differed from system to system.
- It was difficult to compare entities.
- It was difficult to verify the data.
- Entering data only in the IS was inadequate for verifying data.

Therefore, the MFCR decided to use support information systems. The information on individuals, entities, etc. is stored in these systems without dependence on the FCIS. Therefore, they are able to help with checking out the data, which are to be entered into the FCIS. The basic Czech registers listed below serve the purpose of the support systems:

- **The Basic Register for Addresses and Real Estate** supplies data to the Address Register of the Territorial Identification Register (TIR-ADR), which is operated by the Ministry of Labour and Social Affairs, and the Information System of the Real Estate Cadastre (REC), which is operated by the Czech Office for Surveying, Mapping, and Cadastre.
- The **Basic Register of Inhabitants** supplies data to the Central Register of Inhabitants (CRI) that is operated by the Ministry of the Interior.
- The **Basic Register of Enterprise Entities** supplies data to the administrative register, the Access to Registers of Economic Subjects/Entities (ARES), which is operated by the Ministry of Finance.

The prerequisite for using the support systems was to interconnect all relevant offices. The network was built and, since its start in early 2001, it has constantly been evolving. Its name is "GOVBONE." It is based on

Index to International News

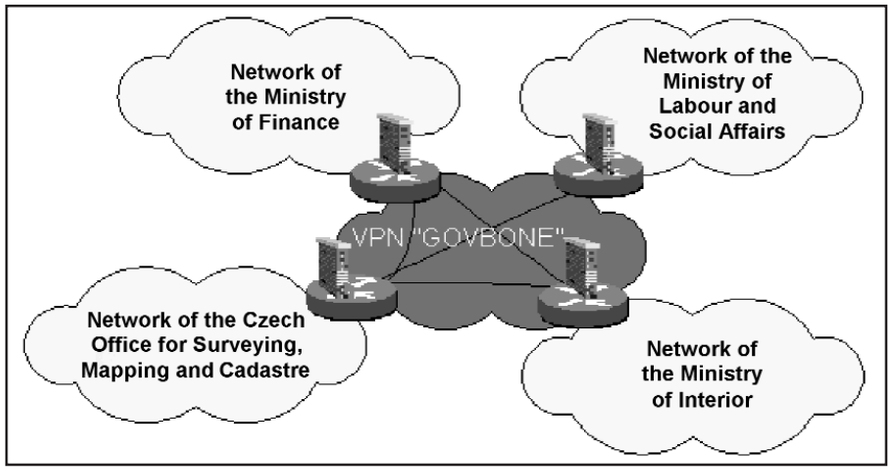


Figure. 1:Interconnection of the Relevant Offices via GOVBONE Network

Internet standards and was designed as a non-public TCP/IP network of the state administration (see Figure 1).

Furthermore, it was necessary to determine the conception of the systems integration. There were two possible alternatives:

- Build one integrated system only.
- Interconnect the information systems already in existence via the communication interface.

The Ministry of Finance chose the second option. Another strategic decision focused on the new standards based on XML. The goal was to build the communication interface that would be of general utility, high scalability, and optional data security. The communication interface is understood to be the independent layer above the network layer.

Furthermore, above the communication interface, a third layer was built. This includes the particular information systems communicating their application programs, as well as their mutual services, with each other. The service is understood to be the precisely determined application function that is defined by its input, its output, and its function logic. Both the input and output are in XML document format. As a client, either the "live" end-user or some other application program can act. The end-user has the special application programs allowing him to use the service (e.g. the Web browser). As to the application program, this can perform some control functions without human participation. Figure 2. illustrates the three-layer principle.

Via the pilot project, the communication interface design was realized and tested. The purpose of that activity was to create relations from the CRSS system to the TIR-ADR and ARES systems in order to improve the CRSS data quality. The required goal has been satisfied. CRSS users can now enter their data very easily. For example, they enter the entity's identification number and press "fill in the form." This brings up the ARES service, which looks up the identification number in ARES and creates the output document containing the business name, the address of the registered office, and other important information on the entity concerned. Such document is sent back to the CRSS application program. In CRSS, the form is now completed. The user confirms the content of the form and then the data are entered into the CRSS database (see Figure 3).

Having confirmed the functionality of the communication interface described above, the principles were generalized and a new project was created. Its name is "Government Interoperability Framework for Financial and Control Systems" (GIF-FCIS). It is planned to extend the services so that all MFCR systems, as well as all their communications with other authorities, would be covered. The MFCR also offered their communication interface to the Czech Office for Public Information

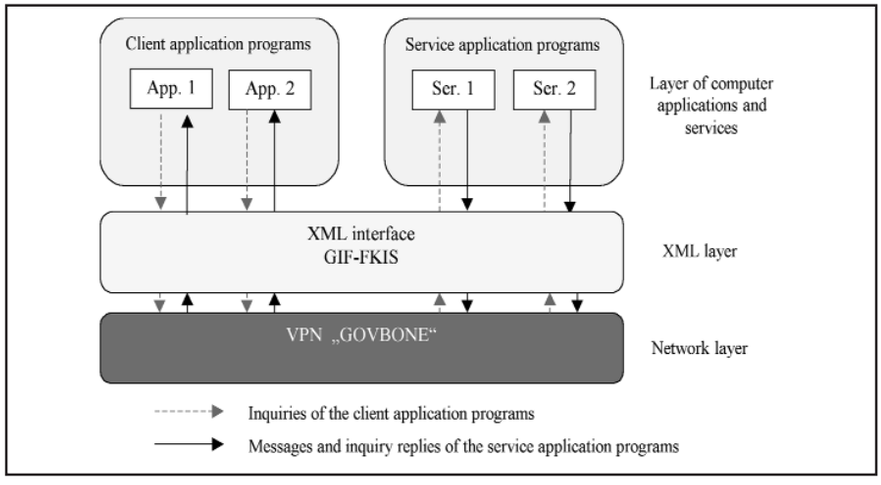


Figure 2: The Three Layers of the Communication Interface

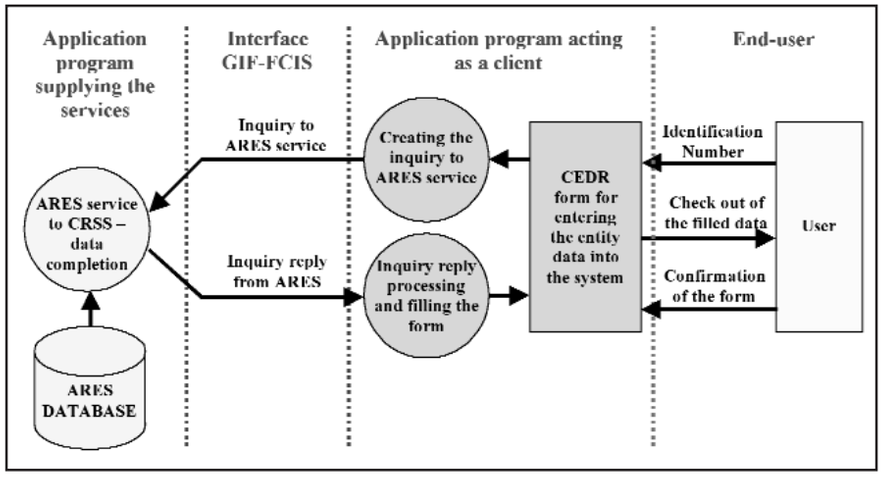


Figure. 3: Schematic View of Using ARES in CRSS

Systems (OPIS) for the purpose of standardization of the communication interface within the framework of the whole Czech state administration.

In conclusion, we present several technical details concerning particular components of the GIF-FCIS. The network layer uses the standard communication protocols available in TCP/IP networks:

- **HTTP/HTTPS** protocols are used for online communication.
- **SMTP** protocol and **MIME/S-MIME** standard are used for off-line communication.
- **FTP/SFPT** protocols are used for communication where large data volumes are transferred.
- **LDAP/LDAPS** protocols are used in the process of authorization and authentication of users.

The XML layer is based on the open international standards that are independent of any concrete technology. This is the question of:

- **XML** as a language for the structured data transmission.
- Standard **XML Schema** for defining the transferred documents.
- Protocol **SOAP** for defining the structured message.
- Standard **BizTalk** for elements needed for the message interchange.
- Standard **XSLT** for operations over the XML documents.
- **Namespaces** standard for the XML schema storage.

The GIF-FCIS security methods are based on the secure version of the communication protocols (HTTPS, S-MIME, SFPT). Furthermore, they use:

- Standard **X.509** for the certificates.
- Standard **XML Signature** for the electronic signature of the XML documents.
- Standard **XML Encryption** for the XML document's encrypting.

For more information, contact Ji_í Roudn_, IT Director of the Ministry of Finance, by telephone at +420 2 2437 3662 or via e-mail at Jiri.Roudny@mfcr.cz, or.

XML Applications to Deliver Government Information and Services to French Citizens and Businesses

By Michel De Rocca
michel.de-rocca@industrie.gouv.fr

The administration has a goal of delivering government information and services to French citizens and businesses using a streamlined, online delivery system. Since mid-2000, the administration has recognized that an interoperability framework information system based on the wide use of XML schemas and DocumentType Definition (DTD) was necessary to achieve this goal. A central repository for XML DTD and schemas was then set up for the benefit of all government bodies (www.atica.pm.gouv.fr/XML). Government bodies now have to submit their schemas for comment before they can be included in this repository. Several administrative areas are already covered, including public libraries, the Supreme Court, Ministry of Equipment regulations, et cetera.

In the meantime, the government portal team (www.service-public.fr) has launched a project now in the implementation phase for municipal Web sites to easily access and integrate (by co-marking techniques, html pages with JavaScript code) information coming from the government portal. For example, national rules from the government portal concerning the voting process could be linked to local administrative offices that deliver the voting cards.

Another application, available since April 2001, concerns the national business register, which has five million registered companies. A secure XML-based application using e-certificates was set up to facilitate access to and use of the central business register by local registration authorities. Each newly established company has to be registered, and significant company events must appear as well on the Sirene register (www.insee.fr).

Also of interest is the initiative of the national organization of accounting companies (Ordre National des Experts Comptables), which encompasses at least 500,000 enterprises, to build a set of XML-based applications for online fiscal, social, and statistical declarations. The Ministry of Finance and Industry is sponsoring the project. Its aim is to give specifically to small enterprises a single interface between their own information systems and the administration's requests for information. The project is ambitious and requires close cooperation between a number of government bodies.

For more information, contact jsaphores@experts-comptables.org.

The Singapore Government – Beyond Traditional Electronic Service Delivery

By Tan Gek Khiang
Senior Consultant
Electronic Services Division,
Government Chief Information Office
InfoComm Development of Singapore (IDA)

Background

The government of Singapore has always been at the forefront in becoming an e-government, allowing businesses and members of the public to easily reach, communicate with, and interact with government agencies. In fact, the Singapore government aims to be a leader in e-government to better serve the nation in the digital economy. To this end, the Singapore government has set aside $1.5 billion for InfoComm technology initiatives in the public sector up to 2003.

Next Wave of Electronic Service Delivery: Integrated Service Delivery

The Singapore government realizes that the success of e-services delivery goes beyond the traditional publish and transact paradigm. Integration of services is the key success factor in delivering government services to citizens. In other words, it is important that government services are streamlined, thus simplifying the lives of citizens with customer-centric, integrated, one-stop e-services.

Such integration of e-services can improve operational efficiency, reduce operational costs, and give citizens a common, unified face of government, thus simplifying their transactions with multiple government agencies.

Public Service Infrastructure: Platform for Integrated Service Delivery

The Singapore government developed the Public Service Infrastructure (PSi) that provides a platform for government agencies and their appointed vendors to rapidly develop and deploy robust Internet applications efficiently and securely. It also provides re-usable components and services, such as electronic payment services, electronic data exchange services, security services, etc.

Improved Interoperability with XML

One key challenge that PSi wants to address is access to agencies' disparate backend databases. PSi allows linking to disparate data sources in government agencies via the Electronic Data Exchange (EDX) Gateway. The EDX Gateway acts as a secure proxy between the e-services on PSi and the government agencies' backend databases. See Figure 1.

Existing agencies have invested a lot in their legacy backend systems, making integration/interoperability within the government an often difficult and expensive process. XML is the key to linking up these disparate data sources because XML, being platform independent and an open standard, is an appropriate messaging format to enable interoperability between heterogeneous platforms.

Improved Portability with XML

XML being platform and language independent is an ideal format for data persistency. In this arena, PSi also uses XML extensively. For example, e-services (i.e. Internet

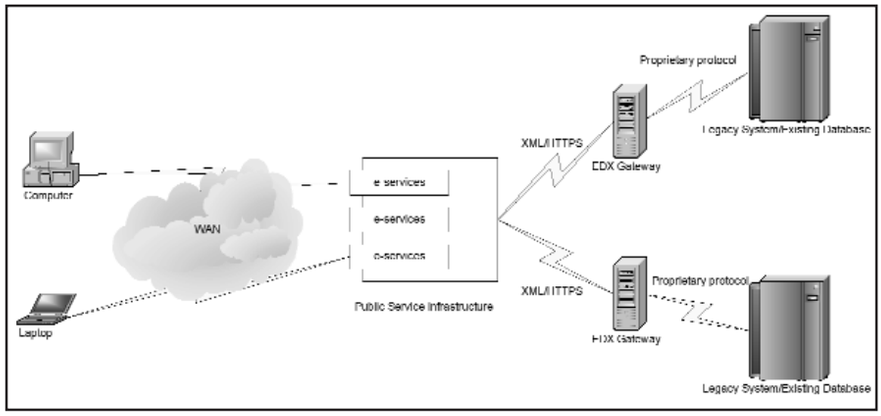


Figure 1. Interoperability with Heterogeneous Platforms via EDX Gateway

applications) on PSi are stored in XML format. This frees PSi from being bound to a particular platform or programming language. In this world of constantly changing and evolving technologies, portability of systems across platforms and languages becomes very crucial in making sure the investments are protected.

Future Plans

The Singapore government is continuing to explore other innovative ways to provide better services to its citizens. The extensibility and portability of XML opens up new opportunities for services to be delivered to citizens.

Improved E-Services Accessibility with XML

One interesting potential use of XML is to provide support for pervasive computing clients such as Personal Digital Assistants (PDAs), mobile phones, and other mobile devices to access e-services.

Presentation and data can be segregated with XML. In the same way, presentation of e-services and its data can be segregated. An e-service can be readily accessible via either a PDA or a mobile phone. This will improve the accessibility of e-services to citizens.

Improved Reusability of Common Components and Services with XML

One benefit that PSi might offer to government agencies is a common service for electronic payment and security authentication. Emerging standards like Simple Object Access Protocol (SOAP) could possibly be the interfacing protocol to these common services, making them accessible over the Internet by advertising via Universal Description, Discovery and Integration (UDDI). Other possible areas for exploration on XML applications include enterprise search and retrieval of XML documents and enterprise portal infrastructure.

For more information, contact Tan Gek Khiang, by telephone at (65) 211-0888 or via e-mail at Tan_Gek_Khiang@ida.gov.sg.

The Architecture of Computer Applications With the XML Publishing Tool "Cocoon" in a Swiss Government Office

By Jean-Bernard Ballif and Hansruedi Vock
Ministry for Informatics and Telecommunications (Bundesamt für Informatik und Telekommunikation-BIT)
Switzerland

The Swiss Federal Office of Information Technology, Systems and Telecommunication, is the main IT provider for the Swiss federal administration. We develop and host a large number of government applications in very different fields. Today, these applications run in many different environments with different technologies. The purpose of the BANAR[1] project is to provide a clear and uniform overview of the architecture of each application, to make the evaluation of our infrastructure easier, and to help our IT planning team make pertinent choices for further development strategies. The project consists of two main parts: the collection of architecture data of all applications and the presentation of these data within a Web application. The project is still under development.

The collected data contain the following information:

- General description of the application (name of application, purpose, kind of architecture, operating system, security, costs, number of users, clients, developers, etc.)

- Components of the application (type, number, operating system, hardware, communication, etc.)
- Relationships with other applications
- Development tools (software, programming languages, operating system, etc.)
- A graphical representation of the application with its components

Some of this information, of course, already exists in one or more forms. Our first task was to collect the existing information and organize it in a rigorous and structured way. For example, we had to extract data from old databases or portfolios. When information was not found, we invited project leaders to fill out very detailed forms. In some cases, we proceeded to interviews. The second task was to write an application to display this information.

There were several primary requirements for the realization of this project:

- The solution had to use simple tools (if possible, open-source products).
- It had to be a Web application running on the government intranet.
- It had to contain a search engine (for example, search for development unit, architecture type, security, component type, operating system, developing tool, programming language, etc.)
- It needed to be able to graphically display statistics (about data bases, programming languages, architecture types, operating systems, etc.). The graphics had to be capable of displaying the current status of the database, so that statistics could be calculated directly from the database and the images processed online.
- It needed to be able to display results in two national languages. (The user should be able to see either the French or the German text and should be able to switch to the other language at any time).
- The application needed to contain a Web form to collect data about new applications. Project leaders should be able to enter information about their applications online, possibly with attachments (images, documentation, etc.). Whether these data are stored in the database immediately or following some verification in the productive environment has not yet been decided.

We were not interested in solutions that mix content and presentation, like CGI, JSP, or servlets. XML appeared to be a very suitable way to achieve our goals. We decided to use the following development tools:

- An MS-Access database to model and collect data.
- A MySQL database to store the data used by the application[2] .
- An Apache Web-server[3].
- A Tomcat servlet container[4]
- The Cocoon XML publishing framework[5] (version 1.8.2)

Our development environments are:

- Windows 2000 or NT to collect and organize the data[6]
- Debian Linux with potato distribution to develop and run the application

Chronologically, the data are first collected in an MS-Access database. When enough data have been collected, this database is converted into a MySQL database with a small Java application. We found it more convenient to keep the MS-Access database as the first storage element.

The Cocoon framework is the main element in this project. Cocoon is implemented as a servlet and should be used with the servlet engine Tomcat. The Apache Web server offers more stability in a productive environment, but is not necessary in a development environment.

The dynamic Web content is generated on request with Cocoon's XSP technology (eXtensible Server Pages). For example, the language information is passed to the XML page as a parameter and also included in a XML attribute:

- <text Language="fr">Text in French</text>
- <text Language="ge">Text in German</text>

The stylesheet is completely independent of the language. The stylesheet only has to check the attribute "Language" against the parameter to the XML page, so that only the selected language is displayed.

For database requests, the Cocoon's SQL Processor performs SQL queries and translates the result set into an XML fragment. This fragment is then inserted in the XML documents. Thus each Web page in BANAR consists of an XML document (including XSP-instructions and SQL tags) containing the required text in French and German and an XSL style sheet to display the data in the selected language.

Up to now, we have developed a few other applications with Cocoon 1 in the same style. Our experiences with XML and Cocoon are very encouraging. The main positive point is that Cocoon's XSP offers a clear separation of content and presentation. Since XSP is processed before any other style sheet, the advantage of XML/XSL remains (separation of content and style). If necessary, it is also possible with XSP to separate logic from content. Cocoon is easy to learn, and the maintenance of Cocoon applications is also very simple. However, the response time

of applications running with Cocoon is a little disappointing, and we are not quite satisfied with the Cocoon documentation. We will consider these points before developing further applications with Cocoon 2.

For more information, contact Hansruedi Vock, Bundesamt für Informatik und Telekommunikation BIT, Monbijoustrasse 74, CH-3003 Bern, by telephone at ++41 31 32 59095 or via e-mail at hansruedi.vock@bit.admin.ch or Jean-Bernard Ballif of the same address, by telephone at ++41 31 32 58079 or via e-mail at jean-bernard.ballif@bit.admin.ch

[1] Bit ANwendungs ARchitektur (German abbreviation)

[2] http://www.mysql.com/

[3] http://httpd.apache.org/

[4] http://jakarta.apache.org/tomcat/index.html

[5] http://xml.apache.org/cocoon/index.html

[6] http://www.microsoft.com/

[7] http://www.debian.org

Government News Exchange via XML

By Asis Unyapoth, Piya Tanthawichian, and Thaweesak Koanantakool
Government Information Technology Services
National Electronics and Computer Technology Center
Ministry of Science Technology and Environment
Government of Thailand

Background

News and up-to-date information are vital in government administration. Realizing this, Thai government offices try to keep their employees informed of relevant news. The Ministry of Science, for example, regularly circulates news from its subdivisions and departments, as well as science and technology news from newspapers and other media. In the past, publication on paper was the only means of circulation. Now, encouraged by the Thai e-government policy, the Internet has been increasingly exploited by the government. Posting news and announcements on the Internet not only reduces publication costs, but also gives members of the public access to government news and announcements.

Nevertheless, there is no easy way to gather relevant news. Typically, each government office employs some human readers for scanning news sources for related information and posting such information on the office Web site. This approach is clearly inefficient and expensive. In this article, we shall discuss an XML technology that supplies government Web sites with the latest news from both government news sources and online newspapers.

The Model

The technology developed is based on the model of establishing a **news exchange center**. News can be dynamically obtained from the news center via Web interfaces, by specifying categories, sources, keywords, dates, and so on. The news obtained can then be processed and put on Web sites. News retrieval and processing can be done each time the Web sites are loaded, ensuring that Web site news is up-to-date.

Since the news center deals with a large number of sources, news published by these sources must be in the same format, otherwise, the government would need to maintain a different program for each source. This is where XML technology fits in. From the news classification and analysis point of view, XML provides an unambiguous way of identifying information. Each piece of news can be clearly marked up—indicating the headline, published date, source, category, and keyword, as shown below.

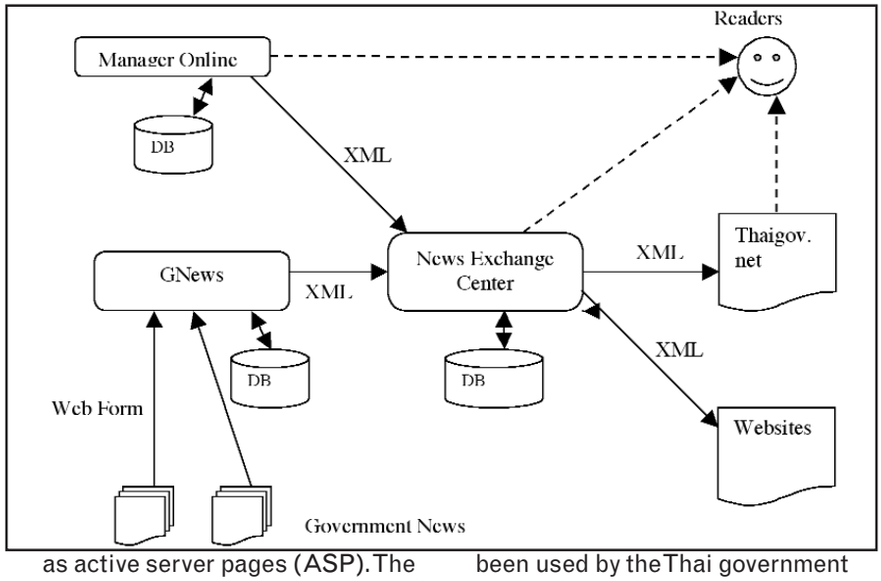
```
<news>
<title>Army kicks off new smart ID cards</title>
<link>http://www.msnbc.com/news/648418.asp?0si=-&cp1=1</link>
<pubdate>10/29/2001</pubdate>
<source>CNN</source>
<sourcelink>http://cnn.com</sourcelink>
<category>Technology</category>
</news>
```

News distributed from the news center to Web sites also uses this XML format. One advantage is that XML separates presentation from information. When a Web site obtains the XML news, it can transform the XML format (using XSL or other techniques) into HTML format, with the style (color, font, etc.) tailored for that particular Web site.

Note that the full content of news is not included in the above format. The news distributed using this format only contains information sufficient for readers to determine what news is of interest to them. If readers wish to pursue the content, they can click on the Web link and read the entire story from the news sources.

The Prototype

A prototype of the news exchange center has been implemented. In cooperation with *Manager Online*, a Thai broadsheet newspaper, the news center is constantly supplied with a wide range of news, from breaking news to entertainment and columnist news. Manager Online (and most online newspapers) uses a database to store details of news—news pages are constructed dynamically using technology such



as active server pages (ASP). The news source, therefore, has no difficulty in providing news for the news center, since news could be easily retrieved and presented in the XML format.

In our experience, most government offices use ad hoc approaches for publishing their news on the Web, typically by creating an HTML for each piece of news and adding a link to it on the main Web site. In this case, producing news in the XML format could be seen as extra work for the Web masters. One of our services, *GNews*, relieves Web masters of these tedious tasks. Registered Web masters (currently limited to government agents) are able to post news and announcements via a Web form. The details of the news posted will then be stored in a database and accessible via active server pages. Recently, we extended *GNews* service so that when a piece of news is submitted to *GNews*, it is also uploaded to the news exchange center—and thus will instantly appear on Web sites that register their interest for such news.

News retrieved from these sources can be read directly from the project Web site, *thaisarn.com*. The technique of dynamic news retrieval has been used by the Thai government portal site, thaigov.net to keep the Web site up-to-date.

Conclusions and Future Work

The news exchange center project is still in its infancy. Nevertheless, many Web sites and news sources have expressed interest in the project. News sources are keen to cooperate in providing news in the XML format, because it potentially increases access to their Web sites. Web masters are keen to use this technology because it keeps the news on their Web sites up-to-date. The main obstacle is technical: different systems use different ways of dealing with XML. Advice to Web masters and news sources has to be given on a case-by-case basis.

On the technical side, the prototype described still requires further improvements, for example, an ability to deal with complex news retrieval. We are also interested in imposing some secure infrastructure that guarantees the confidentiality of news intended for specific groups of readers. Internal news, for example, might be intended for circulation only within an organization and require some authentication mechanism.

Related Works and Technology

Some of the ideas originated from the use of Resource Description File (RDF) and RSS formats (which are dialects of XML) for distributing news, as exploited by Web sites such as *my.userland.com*. The International Press Telecommunications Council (IPTC) also proposes NewsML, an XML standard for news publication. These formats are still under development and are yet to become standard. We reject the RSS format because it provides insufficient tags—none for identifying date, category, and keywords. The NewsML format is also rejected because it is intended for representation of news content, taking into consideration multimedia copyright information, which makes the format highly complex.

For more information, contact Asis Unyapoth by telephone at+ 662-642-7077 ext. 8310 or via e-mail at asis@gits.net.th; Piya Tanthawichian via e-mail at piya@gits.net.th; or Thaweesak Koanantakool via e-mail at htk@nectec.or.th.

XML in the UK Government

By John Borras
Assistant Director, Interoperability and Metadata
Office of the e-Envoy, Cabinet Office
United Kingdom

The government of the United Kingdom (UK) has recognised that Information Systems have the potential to transform government and the services it provides to the public. The UK aims to have all government services online by 2005.

The UK online Citizen Portal is the first step towards establishing a 24-hour point of access to available

government online information and services. In addition to providing access to government services, the site also provides information in the form of "life episodes," such as 'Having a Baby'. The latest life episode to be added (on 11th July 2001) is "Pensions and Retirement." The portal can be accessed at http://www.ukonline.gov.uk/.

From government Web sites, one can:

- Order and pay for company financial returns.
- Check to see if hospital treatment is needed.
- Find out if it is safe to visit Fiji.
- Apply for a passport (and find out what the weather there is going to be).
- Fill out a tax form.
- Keep up-to-date with what is happening in Parliament.
- Read the proceedings in the Commons or Lords.
- Obtain recent legislation.
- Find your MP (and send a fax from your computer to him/her).

The Court Service can provide a range of legal information, including case law and access to online legal advice. And one can now report minor crimes online. Future projects will also allow local government to provide services online.

Underpinning these services and interactions between government departments is the e-Government Interoperability Framework. (See http://www.govtalk.gov.uk/interoperability/egif_document.asp?docnum=363) The Framework prescribes mandatory technical standards and policies. These standards allow information to flow seamlessly across the public sector and provide citizens and businesses with better access to government services. They also set the architecture for UK Online and the Government Gateway (see http://www.gateway.gov.uk/)

The UK has, for example, mandated the W3C standard on XML schemas for use across the UK public sector. By adopting Internet and World Wide Web standards, the Framework aligns government with the rest of industry and serves as a basis for reducing the costs and risks associated with carrying out major IT projects.

Reference material, a top level Information Architecture, a reference model for XML messages, a Data Standards Catalogue, XML Schema Design Guidelines, and a Framework for Metadata supplement these standards. These will soon be online as XML documents. In more detail, these are:

- The top-level information architecture aims to model the key components of all government information needed for interoperability.
- The message reference model is a template for developers to use while designing XML documents for data transactions.
- The Data Standards Catalogue specifies corporate data elements.
- The Schema Design Guidelines lay down best practices for building XML schema that will need to operate within the UK Online space.
- The MetaData Framework is based on Dublin Core and when implemented enhances the searching of reference material.

The Office of the e-Envoy within the Cabinet Office provides the long-term management of the Framework. The Office acts as the focal point for coordinating interoperability efforts throughout government and cooperation with other governments and international bodies.

Specifying policies and standards is not enough. Successful implementation means the provision of support, best practice guidance, XML toolkits, and a repository of centrally agreed schemas.

The UK GovTalk™ Web site (www.govtalk.gov.uk) allows the checking of the very latest public sector specifications. The public can comment on these and also provide us with proposals for solving some of our generic IT problems. The Web site has a repository of government-approved XML schemas, allowing new projects to re-use existing schemas. The Web site allows discussion forums to be set up and managed, and provides best practices and case studies.

The subject range of schemas developed includes filing of tax returns, property management, agricultural subsidy claims, and BS7666 Addresses. Further work is going on in many areas and in due course all those schemas will be made available on GovTalk.

A functional specification for a suite of XML software toolkits required by government agencies is in preparation. These toolkits will provide tangible support for developers of XML-based applications.

In addition to our work within the UK, the Office of the e-Envoy is also working on a European Commission project (MIReG) to establish a European Metadata Framework. (see http://europa.eu.int/ISPO/ida/). One of the outputs from this work will be a set of XML schemas to support the Framework.

For more information, contact John Borras by telephone at +44 20 7276 310 or via e-mail at johnborras@cabinet-office.x.gsi.gov.uk

XML in Australia

By Peter Hallams
Manager Integrated Services
National Office for the Information Economy
Australia

Australia is currently ranked among the leading group of e-government nations by international consulting groups such as accenture.

Federal, state, and local governments are working hard to develop new approaches for delivering information and transactions online. Two innovative projects highlight how the use of eXtensible Markup Language (XML) is now making life easier for people and government in Australia.

Australia Goes Fishing Online

FishOnline (www.fishonline.tas.gov.au) is a recreational fishing 'service pack' that integrates information from multiple government Web sites into a single location. The aim is that a customer—someone who wants to go fishing in Tasmania, the island state off the southeastern coast of Australia—should be able to find information from one interface, rather than following links around the Web. FishOnline is funded by the federal government under the Trials of Innovative Government Online Regional Services (TIGERS) program being conducted in Tasmania.

The Tasmanian company, Dytech Solutions, was chosen as contractor for this project on the basis of their XML development experience demonstrated in Service Tasmania Online (http://www.service.tas.gov.au). Conceived as a service in its own right, FishOnline also draws on existing Web site data and presents it as an integrated whole.

An expandable and editable map interface draws together information from four participating government agencies. It provides the customer with all information that might be relevant about fishing at a particular spot. Agencies can add geographical locations (fishing spots) and populate them with categorized data from their Web sites or with information created specifically for the service pack.

The problem of integrating the four Web site hierarchies to provide better customer service was in itself a significant task. The service pack reduces the need for ongoing maintenance by automatically updating itself on a configurable schedule, which is carried out quickly and easily across the Internet. When data from contributing Web sites is harvested, a parsing tool identifies the content that is required and strips away extraneous formatting information.

Content is currently identified by custom tags inserted in the originating service packs. However, the need for this is being removed in phase two by using unique text and/or tag elements to identify start and finish points for inclusion. The result will be the capacity to include information drawn from most Web sites without having to alter the raw pages/data. Furthermore, where Web sites use standard formatting features or style sheets across the entire site, selection of the appropriate data will become even easier.

The identified content is stored in XML data structures in the Windows file system. The resources are indexed—using the standard indexing technology in Windows 2000, combined with Dytech's custom XML filtering technology—to yield a powerful freeform hierarchical data storage mechanism. Phase two will see this migrated to an XML relational data structure, utilizing Dytech's custom query language support. This will maintain the power of the freeform XML data structures while accessing the power of a relational database.

The service pack is modular and while the technology has been used in this instance for FishOnline, the capacity exists to reuse the central components for many other purposes. Significant effort has gone into ensuring the flexibility and generic features of the framework to make the system applicable for multiple purposes.

Importing Weather Information from the Bureau of Meteorology

The FshOnline service pack also makes use of XML as a transport mechanism for the import of data—tested by importing weather information from the Australian Bureau of Meteorology. In the past, weather information has been extracted from the Bureau's multiple databases and presented in HTML pages on their Web site.

For the FishOnline service pack, the relevant weather information had to be identified, extracted, transported, and then used in several different ways. The structure of data on the Bureau's Web site was varied and complex, and the challenge was to build an XML schema that would be re-usable for multiple purposes and also cope with the vast range of weather information available.

An XML extraction engine was developed within the Bureau to find the required data and create XML files every 10 minutes. The resulting XML files are made publicly available on their Web server, and the service pack harvests them on its own schedule. The service pack can then extract the data it requires and present it in any combination.

A clear benefit of this approach is that all the data related to a particu-

lar fishing spot or area can be brought together in real time on one page, formatted as required. This is a feature not found as yet on the Bureau's own Web site. The capacity exists for the Bureau to re-use this output by re-extracting the presented pages for its own site.

Although this particular use of XML is only in its infancy, the possibilities for the Bureau to slice and dice their data for multiple purposes, in collaboration with others, are limited only by the imagination.

XML in the Nation's Capital

Located in Canberra, Australia's national capital, the Australian Capital Territory (ACT) Government is recognized as a world leader in e-government services. The ACT Government provides both state and local government services and works with the Information and Communications Technology (ICT) industry and the Commonwealth Government to develop innovative e-government solutions.

Recently the ACT Government extended the long-term relationship it has with Wizard Information Services for the provision of such solutions. ACT Government Agencies have been using Wizard's XML based e-business solution Meta*WizDom and their e-Business Readiness methodology to rapidly deploy portals and content rich Web sites.

The government has chosen a new whole-of-government vision for e-government. This project replaces some 160 separate Internet and intranet systems and will also deliver ICT industry development opportunities to small and medium enterprises (SMEs) in e-business, using eXtensible Markup Language (XML) and eXtensible Stylesheet Language (XSL) technologies.

The ACT Government's vision is for a single source of information that is topic-based and delivered uniformly through any number of delivery channels including the Internet, call centers, Web TV, kiosks, and shop fronts. A significant element of this vision is that it overcomes the need for users to understand the nature and structure of government in order to gain effective access to public information and services.

The ACT Government is taking a world leading approach to resolving these and other related issues. Meta*WizDom is already used by the ACT Government to build a number of innovative solutions including the Canberra Connect portal www.canberraconnect.act.gov.au. Launched in February 2001, it is republished on an almost daily basis and provides people and businesses with a constant flow of up-to date information.

Meta*WizDom allows the development of related portals and 'umbrella portals' that provide an entry point to large and complex organizations, allowing searching across all related sites and managing links and content between the sites. The content in sites such as Canberra Connect can be maintained directly by contributing government agencies that can update or create completely new information and submit it for publication within minutes.

Other sites provide examples of the flexibility, increased connectivity, and ease of use that are characteristics of this approach—such as the Department of Urban Services portal (www.urbanservices.act.gov.au) and Environment ACT Web site (www.environment.act.gov.au).

Another innovative e-government project, Communities Online (www.actcommunitiesonline.org), is the first phase of a very ambitious program. It enables community organizations and groups to publish their own Web sites without the need for any technical knowledge. This service is provided free to all community organizations and groups, with the provision that content must be 'family friendly.' This partnership between the ACT Government and Meta*WizDom provides one of the most advanced e-government solutions to further demonstrate the benefits of XML.

For more information, contact Peter Hallams via e-mail at Peter.Hallams@ogo.gov.au.

U.S. General Services Administration
Office of Governmentwide Policy
Office of Intergovernmental Solutions